# (19) World Intellectual Property Organization International Bureau





# (43) International Publication Date 7 February 2002 (07.02.2002)

# **PCT**

# (10) International Publication Number WO 02/10382 A2

- (51) International Patent Classification<sup>7</sup>: C12N 15/12, 15/11, 9/00, C07K 14/47, C12Q 1/68, G01N 33/577, A61K 31/713
- (21) International Application Number: PCT/EP01/08309
- (22) International Filing Date: 18 July 2001 (18.07.2001)
- (25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

60/221,513

28 July 2000 (28.07.2000) U

(71) Applicant and

- (72) Inventor: WISSENBACH, Ulrich [DE/DE]; Institut für Pharmakologie und Toxikologie der Uni, versität des Saarlandes, 66421 Homburg (DE).
- (74) Agent: HUBER, Bernard; Huber & Schüssler, Truderinger Str. 246, 81825 München (DE).

- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, Cl, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

#### Published:

 without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

0382 A

(54) Title: TRP8, TRP9 AND TRP10, NOVEL MARKERS FOR CANCER

(57) Abstract: The present invention relates to gene expression in normal cells and cells of malignant tumors and particularly to novel markers associated with cancer, Trp8, Trp9 and Trp10, and the genes encoding Trp8, Trp9 and Trp10. Also provided are vectors, host cells, antibodies, and recombinant methods for producing these human proteins. The invention further relates to diagnostic and therapeutic methods useful for diagnosing and treating a tumor.

# Trp8, Trp9 and Trp10, novel markers for cancer

# FIELD OF THE INVENTION

The present invention relates to gene expression in normal cells and cells of malignant tumors and particularly to novel markers associated with cancer, Trp8, Trp9 and Trp10, and the genes encoding Trp8, Trp9 and Trp10

# **BACKROUND OF THE TECHNOLOGY**

Prostate cancer is one of the most common diseases of older men world wide. Diagnosis and monitoring of prostate cancer is difficult because of the heterogeneity of the disease. For diagnosis different grades of malignancy can be distinguished according to the Gleason-Score Diagnosis. For this diagnosis a prostate tissue sample is taken from the patient by biopsy and the morphology of the tissue is investigated. However, this approach only yields subjective results depending on the experience of the pathologist. For confirmation of these results and for obtaining an early diagnosis an additional diagnostic method can be applied which is based on the detection of a prostate specific antigen (PSA). PSA is assayed in serum samples, blood samples etc. using an anti-PSA-antibody. However, since in principle PSA is also expressed in normal prostate tissue there is a requirement for the definition of a threshold value (about 4 ng/ml PSA) in order to be able to distinguish between normal and malign prostate tissue. Unfortunately, this diagnostic method is quite insensitive and often yields false-positive results. Moreover, by using this diagnostic method any conclusions as regards the grade of malignancy, the progression of the tumor and its potential for metastasizing cannot be drawn. Thus, the use of molecular markers would be helpful to distinguish benign from malign tissue and for grading and staging prostate carcinoma, particularly for patients with metastasizing prostate cancer having a very bad prognosis.

The above discussed limitations and failings of the prior art to provide meaningful specific markers which correlate with the presence of prostate tumors, in particular metastasizing tumors, has created a need for markers which can be used diagnostically, prognostically and therapeutically over the course of this disease. The present invention fulfils such a need by the provision of Tpr8, Trp9 and Trp10 and the genes encoding Trp8, Trp9 and Trp10: The genes encoding Trp8 and Trp10 are expressed in prostate carcinoma and prostatic metastasis, but

not in normal prostate, benign hyperplasia (BHP) and intraepithelial prostatic neoplasia (PIN). Furthermore, expression of Trp10 transcripts is detectable in carcinoma but not in healthy tissue of the lung, the prostate, the placenta and in melanoma.

# SUMMARY OF THE INVENTION

The present invention is based on the isolation of genes encoding novel markers associated with a cancer, Trp8, Trp9 and Trp10. The new calcium channel proteins Trp8, Trp9 and Trp10 are members of the trp (transient receptor potential) - family, isolated from human placenta (Trp8a and Trp8b) and humane prostate (Trp9, Trp10a and Trp10b). Trp proteins belong to a steadily growing family of Ca<sup>2+</sup> selective and non selective ion channels. In the recent years seven Trp proteins (trp1 - trp7) have been identified and suggested to be involved in cation entry, receptor operated calcium entry and pheromone sensory signaling. Structurally related to the trp proteins are the vanilloid receptor (VR1) and the vanilloid like receptor (VRL-1) both involved in nociception triggered by heat. Furthermore, two calcium permeable channels were identified in rat small intestine (CaT1) and rabbit kidney (ECaC). These distantly related channels are suggested to be involved in the uptake of calcium ions from the lumen of the small intestine (CaT1) or in the reuptake of calcium ions in the distal tubule of the kidney (ECaC). Common features or the Trp and related channels are a proposed structure comprising six transmembrane domains including several conserved amino acid motifs. In the present invention the cloning and expression of a CaT1 like calcium channel (Trp8) from human placenta as well as Trp9 and Trp10 (two variants, Trp10a and Trp10b) is described. Two polymorphic variants of the Trp8 cDNA were isolated from placenta (Trp8a and Trp8b). Transient expression of the Trp8b cDNA in HEK (human embryonic kidney) cells results in cytosolic calcium overload implicating that the Trp8 channel is constitutive open in the expression system. Trp8 induces highly calcium selective inward currents in HEK cells. The C-terminus of the Trp8 protein binds calmodulin in a calcium dependent manner. The Trp9 channel is expressed in trophoblasts and syncytiotrophoblasts of placenta and in pancreatic acinar cells. Furthermore, the Trp8 channel is expressed in prostatic carcinoma and prostatic metastases, but not in normal tissue of the prostate. No expression of Trp8 transcripts is detectable in benign prostatic hyperplasia (BPH) or prostatic intraepithelial neoplasia (PIN). Therefore, the Trp8 channel is exclusively expressed in malign prostatic tissues and serves as molecular marker for prostate cancer. From the experimental results it is also apparent that the

2

modulation of Trp8 and/or Trp10, e.g. the inhibition of expression or activity, is of therapeutic interest, e.g. for the prevention of tumor progression.

The present invention, thus, provides a Trp8, Trp9 and Trp10 protein, respectively, as well as nucleic acid molecule encoding the protein and, moreover, an antisense RNA, a ribozyme and an inhibitor, which allow to inhibit the expression or the activity of Trp8, Trp9 and/or Trp10.

In one embodiment, the present invention provides a diagnostic method for detecting a prostate cancer or endometrial cancer (cancer of the uterus) associated with Trp8 or Trp10 in a tissue of a subject, comprising contacting a sample containing Trp8 and/or Trp10 encoding mRNA with a reagent which detects Trp8 and/or Trp10 or the corresponding mRNA.

In a further embodiment, the present invention provides a diagnostic method for detecting a melanoma, chorion carcinoma, cancer of the lung and of the prostate in a tissue of a subject, comprising contacting a sample with a reagent which detects Trp10a and/or Trp10b antisense transcripts or Trp10a and/or Trp10b related antisense transcripts.

In another embodiment, the present invention provides a method of treating a prostate tumor, carcinoma of the lung, carcinoma of the placenta (chorion carcinoma) or melanoma associated with Trp8 and/or Trp10, comprising administering to a subject with such an disorder a therapeutically effect amount of a reagent which modulates, e.g. inhibits, expression of Trp8 and/or Trp10 or the activity of the protein, e.g. the above described compounds.

Finally, the present invention provides a method of gene therapy comprising introducing into cells of a subject an expression vector comprising a nucleotide sequence encoding the above mentioned antisense RNA or ribozyme, in operable linkage with a promoter.

#### BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1: A, phylogenetic relationship of trp and related proteins. B, hydropathy plot of the Trp8 protein sequence according to Kyte and Doolittle. C, alignment of Trp8a/b to the epithelial calcium channels ECaC (from rabbit) and Vr1 (from rat). Putative transmembrane domains are underlined.

Figure 2: A, polymorphism of the Trp8 gene. The polymorphic variants Trp8a and Trp8b differ in five base pairs resulting in three amino acid exchanges in the derived protein sequences. Specific primers were derived from the Trp8 gene as indicated by arrows. B, the Trp8a and Trp8b genes are distinguishable by a single restriction site. Genomic fragments of the Trp8 gene can be amplified using specific primers (shown in A). The genomic fragment of the Trp8b gene contains an additional site of the restriction enzyme BSP1286I (B). C, the Trp8 gene is located on chromosome 7. D, genotyping of eleven human subjects. A 458 bp genomic fragment of the Trp8 gene was amplified using specific primers (shown in A) and restricted with BSP1286I. The resulting fragments were analyzed by PAGE electrophoresis.

Figure 3: The Trp8b protein is a calcium selective ion channel. A, representative trace of a pdiTrp8b transfected HEK 293 cell. Trp8b mediated currents are activated by voltage ramps (-100 mV - +100 mV) of 100 msec at -40 mV or +70 mV holding potential. 1, Trp8b currents in the presence at 2mm  $[Ca^{2+}]_0$ ; 2, effect of solution switch alone 3, switch to nominal zero calcium solution. B, Trp8b currents in the presence of zero divalent cations. C, current voltage relationship of the currents shown in A. Inset, leak subtracted current. D, current voltage relationship of the current shown in B. E, statistics of representative experiments. Black: Trp8 transfected cells, gray: control cells. Columns from left to right: Trp8 currents at -40 mV (n=12) and +70 mV holding potential (n=12). Trp8 currents in standard bath solution including 120 mM NMDG without sodium (n=7) and with nominal zero calcium ions (n=8) or in the presence of 1mM EGTA with zero divalent cations (n=6). F, representative changes in  $[Ca^{2+}]_i$  in Trp8b transfected HEK cells (gray) and controls (black) in the presence or absence of 1mM  $[Ca^{2+}]_0$ . Inset, relative increase of cytosolic calcium concentration of Trp8b transfected HEK cells, before and after readdition of 1 mM  $[Ca^{2+}]_0$  in comparison to control cells.

Figure 4: The C-terminal region of the Trp8 protein binds calmodulin. A, N- and C-terminal fragments of the Trp8 protein used for calmodulin binding studies. B, the Trp8 protein and a truncated Trp8 protein which was in vitro translated after MunI cut of the cDNA, which lacks the C-terminal 32 amino acid residues, were in vitro translated in the presence of <sup>35</sup>S-methionine and incubated with calmodulin coupled agarose beads in the presence of 1 mM Ca<sup>2+</sup> or 2 mM EGTA. C, calmodulin binding to N- and C-terminal fragments of the Trp8protein in the presence of Ca<sup>2+</sup> (1 mM) or EGTA (2 mM)

Figure 5: Expression pattern of the Trp8 cDNA. A, Northern blots (left panels, Clontech, Palo Alto) were hybridized using a 348 bp NcoI/BamHT fragment of the Trp9 cDNA. The probe hybridizes to mRNA species isolated from the commercial blot, but not to mRNA species isolated from benign prostate hyperplasia (right panel, mRNA isolated from 20 human subjects with benign prostate hyperplasia). B,C, in situ hybridization with biotinylated Trp8 specific oligonucleotides on slides of human tissues. Left column antisense probes, right column sense probes. D, antinsense probes.

<u>Figure 6:</u> Differential expression of Trp8 cDNA in human prostate. A-F, in situ hybridization with prostatic tissues. A, normal prostate, B, primary carcinoma, C, benign hyperplasia, D, rezidive carcinoma, E, prostatic intraepithelial neoplasia, F, lymphnode metastasis of the prostata.

Figure 7: Trp8a cDNA sequence and derived amino acid sequence

Figure 8: A, Trp8b cDNA sequence and derived amino acid sequence

B, cDNA sequence of splice variant 1 (12B1)

C, cDNA sequence of splice variant 2 (17-3)

D, cDNA sequence of splice variant 3 (23A3)

E, cDNA sequence of splice variant 4 (23C3)

Figure 9: A, Trp9 cDNA sequence and derived amino acid sequence B, cDNA sequence of splice variant 15 and derived amino acid sequence.

Figure 10: A, cDNA sequence of Trp10a and derived amino acid sequence, B, cDNA fragment of Trp10a and derived amino acid sequence.

Figure 11: cDNA sequence of Trp10b and derived amino acid sequence.

Figure 12: Expression of Trp8 mRNA in human endometrial cancer or cancer of the uterus. A - D, in situ hybridization with slides of endometrial cancer hybridized with Trp8 antisense (left column) or sense probes as controls (right column). E - F, Trp8 antisense probes hybridized to slides of normal endometrium. It can be clearly seen no hybridization occurs with normal endometrial tissue.

# Figure 13: Expression of human Trp9 and Trp10 genes

Northern blots were hybridized using Trp9 (upper panel) or Trp10 (lower panel) specific probes. Expression of the Trp9 cDNA is detectable in many tissues including human prostate and colon as well as in benign prostatic hyperplasia. Expression of Trp10 cDNA is detectable in human prostate of a commercial northern blot (Clontech, right side). This Northern blot contains prostatic tissue collected from 15 human subjects in the range of 14 - 60 years of age. No expression of Trp10 cDNA was detectable in benign prostatic hyperplasia (left side).

Figure 14: Expression of Trp10 transcripts and Trp10-antisense transcripts in human prostate cancer and metastasis of a melanoma. In situ hybridizations of slides hybridized with Trp10-antisense (A-E, K-N) and Trp10 related sense probes (F-J, P-R). It can clearly be seen that both probes detect the same cancer cells indicating that these cancer cells express Trp10 transcripts as well as Trp10-antisense transcripts. S, no Trp10 expression is detectable in benign hyperplasia of the prostate (BPH). O and T, show expression of Trp10 transcripts (O) and Trp10-antisense transcripts (T) in a metastasis of a melanoma in human lung. Melanoma cancer cells express both Trp10 transcripts and Trp10-antisense transcripts.

# DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to an isolated nucleic acid molecule encoding the human prostate carcinoma associated protein Trp8a, Trp8b, Trp9, Trp10a or Trp10b or a protein exhibiting biological properties of Trp8a, Trp8b, Trp9, Trp10a or Trp10b and being selected from the group consisting of

- (a) a nucleic acid molecule encoding a protein that comprises the amino acid sequence depicted in Figure 7, 8A, 9,10 or 11;
- (b) a nucleic acid molecule comprising the nucleotide sequence depicted in Figure 7, 8A, 9,10, or 11;
- (c) a nucleic acid molecule included in DSMZ Deposit no. DSM 13579 (deposit date: 28 June 2000), DSM 13580 (deposit date: 28 June 2000), DSM 13584 (deposit date: 5 July 2000), DSM 13581 (deposit date: 28 June 2000) or DSM ....(deposit date:....);
- (d) a nucleic acid molecule with hybridizes to a nucleic acid molecule specified in (a) to (c)

(e) a nucleic acid molecule the nucleic acid sequence of which deviates from the nucleic sequences specified in (a) to (d) due to the degeneration of the genetic code, and

(f) a nucleic acid molecule, which represents a fragment, derivative or allelic variation of a nucleic acid sequence specified in (a) to (e).

As used herein, a protein exhibiting biological properties of Trp8a, Trp8b, Trp9,Trp10a or Trp10b is understood to be a protein having at least one of the activities as illustrated in the Examples, below.

As used herein, the term "isolated nucleic acid molecule, includes nucleic acid molecules substantially free of other nucleic acids, proteins, lipids, carbohydrates or other materials with which it is naturally associated.

In a first embodiment, the invention provides an isolated nucleic acid molecule encoding the human prostate carcinoma associated protein Trp8a, Trp8b, Trp9,Trp10a or Trp10b comprising the amino acid sequence depicted in Figure 7, 8A, 9,10 or 11. The present invention also provides a nucleic acid molecule comprising the nucleotide sequence depicted in Figure 7, 8A, 9,10 or 11.

The present invention provides not only the generated nucleotide sequence identified in Figure 7, 8A, 9,10 or 11, respectively and the predicted translated amino acid sequence, respectively, but also plasmid DNA containing a Trp8a cDNA deposited with the DSMZ, under DSM 13579, a Trp8b cDNA deposited with the DSMZ, under DSM 13580, a Trp9 cDNA deposited with the DSMZ, under DSM 13581, and a Trp10b cDNA deposited with the DSMZ, under DSM...., respectively. The nucleotide sequence of each deposited Trp-clone can readily be determined by sequencing the deposited clone in accordance with known methods. The predicted amino acid sequence can then be verified from such deposits. Moreover, the amino acid sequence of the protein encoded by each deposited clone can also be directly determined by peptide sequencing or by expressing the protein in a suitable host cell containing the deposited Trp-encoding DNA, collecting the protein, and determining its sequence.

The nucleic acid molecules of the invention can be both DNA and RNA molecules. Suitable DNA molecules are, for example, genomic or cDNA molecules. It is understood that all

nucleic acid molecules encoding all or a portion of Trp8a, Trp8b, Trp9,Trp10a or Trp10b are also included, as long as they encode a polypeptide with biological activity. The nucleic acid molecules of the invention an be isolated from natural sources or can be synthesized according to know methods.

The present invention also provides nucleic acid molecules which hybridize to the above nucleic acid molecules. As used herein, the term "hybridize,, has the meaning of hybridization under conventional hybridization conditions, preferably under stringent conditions as described, for example, in Sambrook et al., Molecular Cloning, A Laboratory Manual 2nd edition (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. Also contemplated are nucleic acid molecules that hybridize to the Trp nucleic acid molecules at lower stringency hybridization conditions. Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages of formamide result in lowered stringency), salt conditions, or temperature. For example, lower stringency conditions include an overnight incubation at 37°Cin a solution comprising 6X SSPE (20X SSPE = 3M NaCl; 9.2M NaH<sub>2</sub>PO<sub>4</sub>; 0.02M EDTA, pH7.4), 0.5% SDS, 30% formamide, 100 µg/ml salmon sperm blocking DNA, following by washes at 50°C with 1 X SSPE, 0.1% SDS. In addition, to achieve even lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5X SSC). Variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

Nucleic acid molecules that hybridize to the molecules of the invention can be isolated, e.g., from genomic or cDNA libraries that were produced from human cell lines or tissues. In order to identify and isolate such nucleic acid molecules the molecules of the invention or parts of these molecules or the reverse complements of these molecules can be used, for example by means of hybridization according to conventional methods (see, e.g., Sambrook et al., supra). As a hybridization probe nucleic acid molecules can be used, for example, that have exactly or basically the nucleotide sequence depicted in Figure 7, 8A, 9,10 or 11, respectively, or parts of these sequences. The fragments used as hybridization probe can be synthetic

fragments that were produced by means of conventional synthetic methods and the sequence of which basically corresponds to the sequence of a nucleic acid molecule of the invention.

The nucleic acid molecules of the present invention also include molecules with sequences that are degenerate as a result of the genetic code.

In a further embodiment, the present invention provides nucleic acid molecules which comprise fragments, derivatives and allelic variants of the nucleic acid molecules described above encoding a protein of the invention. "Fragments,, are understood to be parts of the nucleic acid molecules that are long enough to encode one of the described proteins. These fragments comprise nucleic acid molecules specifically hybridizing to transcripts of the nucleic acid molecules of the invention. These nucleic acid molecules can be used, for example, as probes or primers in the diagnostic assay and/or kit described below and, preferably, are oligonucleotides having a length of at least 10, in particular of at least 15 and particularly preferred of at least 50 nucleotides. The nucleic acid molecules and oligonucleotides of the invention can also be used, for example, as primers for a PCR reaction. Examples of particular useful probes (primers) are shown in Tables 1 and 2.

#### Table 1

Trp8 probes used for in situ hybridization:

Probes (antisense)

- 1.) 5' TCCGCTGCCGGTTGAGATCTTGCC 3'
- 2.) 5' CTTGCTCCATAGGCAGAGAATTAG 3'
- 3.) 5' ATCCTCAGAGCCCCGGGTGTGGAA3'

# Controls (sense)

- 1.) 5' GGCAAGATCTCAACCGGCAGCGGA 3'
- 2.) 5' CTAATTCTCTGCCTATGGAGCAAG 3'
- 3.) 5' TTCCACACCCGGGGCTCTGAGGAT 3'

# Tabelle 2

Trp10 probes used for the in situ hybridizations shown in Figure 14:

#### Probes (antisense)

1.) 5' GCTTCCACCCCAAGCTTCACAGGAATAGA 3' (Figure 14 A, 14B)

2.) 5' GGCGATGAAATGCTGGTCTGTGGC 3' (Figure 14C, 14D, 14N, 14S, 14O)

3.) 5' ATCTTCCAGTTCTTGGTGTCTCGG 3' (Figure 14E, 14K)

4.) 5' GCTGCAGTACTCCTGCACCAGGAA 3' (Figure 14L, 14M)

# Probes (sense)

1.) 5' TCTATTCCTGTGAAGCTTGGGGTGGAAGC 3' (Figure 14F, 14G)

2.) 5' GCCACAGACCAGCATTTCATCGCC 3' (Figure 14H, 14I, 14T)

3.) 5' CCGAGACACCAAGAACTGGAAGAT 3' (Figure 14J, 14P)

4.) 5' TTCCTGGTGCAGGAGTACTGCAGC 3' (Figure 14Q, 14R)

The term "derivative, in this context means that the sequences of these molecules differ from the sequences of the nucleic acid molecules described above at one or several positions but have a high level of homology to these sequences. Homology hereby means a sequence identity of at least 40%, in particular an identity of at least 60%, preferably of more than 80% and particularly preferred of more than 90%. These proteins encoded by the nucleic acid molecules have a sequence identity to the amino acid sequence depicted in Figure 7, 8A, 9, 10 and 11, respectively, of at least 80%, preferably of 85% and particularly preferred of more than 90%, 97% and 99%. The deviations to the above-described nucleic acid molecules may have been produced by deletion, substitution, insertion or recombination. The definition of the derivatives also includes splice variants, e.g. the splice variants shown in Figures 8B to 8E and 9B.

The nucleic acid molecules that are homologous to the above-described molecules and that represent derivatives of these molecules usually are variations of these molecules that represent modifications having the same biological function. They can be naturally occurring variations, for example sequences from other organisms, or mutations that can either occur naturally or that have been introduced by specific mutagenesis. Furthermore the variations can be synthetically produced sequences. The allelic variants can be either naturally occurring variants or synthetically produced variants or variants produced by recombinant DNA processes.

Generally, by means of conventional molecular biological processes it is possible (see, e.g., Sambrook et al., supra) to introduce different mutations into the nucleic acid molecules of the invention. As a result Trp proteins or Trp related proteins with possibly modified biological properties are synthesized. One possibility is the production of deletion mutants in which nucleic acid molecules are produced by continuous deletions from the 5'- or 3'-terminal of the coding DNA sequence and that lead to the synthesis of proteins that are shortened accordingly. Another possibility is the introduction of single-point mutation at positions where a modification of the amino aid sequence influences, e.g., the ion channel properties or the regulations of the trp-ion channel. By this method muteins can be produced, for example, that possess a modified ion conducting pore, a modified K<sub>m</sub>-value or that are no longer subject to the regulation mechanisms that normally exist in the cell, e.g. with regard to allosteric regulation or covalent modification. Such muteins might also be valuable as therapeutically useful antagonists of Trp8a, Trp8b, Trp9,Trp10a or Trp10b, respectively.

For the manipulation in prokaryotic cells by means of genetic engineering the nucleic acid molecules of the invention or parts of these molecules can be introduced into plasmids allowing a mutagenesis or a modification of a sequence by recombination of DNA sequences. By means of conventional methods (cf. Sambrook et al., supra) bases can be exchanged and natural or synthetic sequences can be added. In order to link the DNA fragments with each other adapters or linkers can be added to the fragments. Furthermore, manipulations can be performed that provide suitable cleavage sites or that remove superfluous DNA or cleavage sites. If insertions, deletions or substitutions are possible, in vitro mutagenesis, primer repair, restriction or ligation can be performed. As analysis method usually sequence analysis, restriction analysis and other biochemical or molecular biological methods are used.

The proteins encoded by the various variants of the nucleic acid molecules of the invention show certain common characteristics, such as ion channel activity, molecular weight, immunological reactivity or conformation or physical properties like the electrophoretical mobility, chromatographic behavior, sedimentation coefficients, solubility, spectroscopic properties, stability; pH optimum, temperature optimum.

The invention furthermore relates to vectors containing the nucleic acid molecules of the invention. Preferably, they are plasmids, cosmids, viruses, bacteriophages and other vectors

usually used in the field of genetic engineering. Vectors suitable for use in the present invention include, but are not limited to the T7-based expression vector for expression in mammalian cells and baculovirus-derived vectors for expression in insect cells. Preferably, the nucleic acid molecule of the invention is operatively linked to the regulatory elements in the recombinant vector of the invention that guarantee the transcription and synthesis of an RNA in prokryotic and/or eukaryotic cells that can be translated. The nucleotide sequence to be transcribed can be operably linked to a promoter like a T7, metallothionein I or polyhedrin promoter.

In a further embodiment, the present invention relates to recombinant host cells transiently or stable containing the nucleic acid molecules or vectors or the invention. A host cell is understood to be an organism that is capable to take up *in vitro* recombinant DNA and, if the case may be, to synthesize the proteins encoded by the nucleic acid molecules of the invention. Preferably, these cells are prokaryotic or eukaryotic cells, for example mammalian cells, bacterial cells, insect cells or yeast cells. The host cells of the invention are preferably characterized by the fact that the introduced nucleic acid molecule of the invention either is heterologous with regard to the transformed cell, i.e. that it does not naturally occur in these cells, or is localized at a place in the genome different from that of the corresponding naturally occurring sequence.

A further embodiment of the invention relates to isolated proteins exhibiting biological properties of the human prostate carcinoma associated protein Trp8a, Trp8b, Trp9,Trp10a or Trp10b and being encoded by the nucleic acid molecules of the invention, as well as to methods for their production, whereby, e.g., a host cell of the invention is cultivated under conditions allowing the synthesis of the protein and the protein is subsequently isolated from the cultivated cells and/or the culture medium. Isolation and purification of the recombinantly produced proteins may be carried out by conventional means including preparative chromatography and affinity and immunological separations involving affinity with an anti-Trp8a-, anti-Trp8b-, anti-Trp9-,anti-Trp10a- or anti-Trp10b-antibody, respectively.

As used herein, the term "isolated protein, includes proteins substantially free of other proteins, nucleic acids, lipids, carbohydrates or other materials with which it is naturally associated. Such proteins however not only comprise recombinantly produced proteins but include isolated naturally occurring proteins, synthetically produced proteins, or proteins

produced by a combination of these methods. Means for preparing such proteins are well understood in the art. The Trp proteins are preferably in a substantially purified form. A recombinantly produced version of a human prostate carcinoma associated protein Trp8a, Trp8b, Trp9,Trp10a or Trp10b protein, including the secreted protein, can be substantially purified by the one-step method described in Smith and Johnson, Gene 67, 31-40 (1988).

In a further preferred embodiment, the present invention relates to an antisense RNA sequence characterised that it is complementary to an mRNA transcribed from a nucleic acid molecule of the present invention or a part thereof and can selectively bind to said mRNA. said sequence being capable of inhibiting the synthesis of the protein encoded by said nucleic acid molecules, and a ribozyme characterised in that it is complementary to an mRNA transcribed from a nucleic acid molecule of the present invention or a part thereof and can selectively bind to and cleave said mRNA, thus inhibiting the synthesis of the proteins encoded by said nucleic acid molecules. Riboyzmes which are composed of a single RNA chain are RNA enzymes, i.e. catalytic RNAs, which can intermolecularly cleave a target RNA, for example the mRNA transcribed from one of the Trp genes. It is now possible to construct ribozymes which are able to cleave the target RNA at a specific site by following the strategies described in the literature. (see, e.g., Tanner et al., in: Antisense Research and Applications, CRC Press Inc. (1993), 415-426). The two main requirements for such ribozymes are the catalytic domain and regions which are complementary to the target RNA and which allow them to bind to its substrate, which is a prerequisite for cleavage. Said complementary sequences, i.e., the antisense RNA or ribozyme, are useful for repression of Trp8a-, Trp8b, Trp9-,Trp10a- and Trp10b-expression, respectively, i.e. in the case of the treatment of a prostate cancer or endometrial cancer (carcinoma of the uterus). Preferably, the antisense RNA and ribozyme of the invention are complementary to the coding region. The person skilled in the art provided with the sequences of the nucleic acid molecules of the present invention will be in a position to produce and utilise the above described antisense RNAs or ribozymes. The region of the antisense RNA and ribozyme, respectively, which shows complementarity to the mRNA transcribed from the nucleic acid molecules of the present invention preferably has a length of at least 10, in particular of at least 15 and particularly preferred of at least 50 nucleotides.

In still a further embodiment, the present invention relates to inhibitors of Trp8a, Trp8b, Trp9, Trp10a and Trp10b, respectively, which fulfill a similar purpose as the antisense RNAs or

ribozymes mentioned above, i.e. reduction or elimination of biologically active Trp8a, Trp8b, Trp9, Trp10a or Trp10b molecules. Such inhibitors can be, for instance, structural analogues of the corresponding protein that act as antagonists. In addition, such inhibitors comprise molecules identified by the use of the recombinantly produced proteins, e.g. the recombinantly produces protein can be used to screen for and identify inhibitors, for example, by exploiting the capability of potential inhibitors to bind to the protein under appropriate conditions. The inhibitors can, for example, be identified by preparing a test mixture wherein the inhibitor candidate is incubated with Trp8a, Trp8b, Trp9, Trp10a or Trp10b, respectively, under appropriate conditions that allow Trp8a, Trp8b, Trp9, Trp10a or Trp10b to be in a native conformation. Such an in vitro test system can be established according to methods well known in the art. Inhibitors can be identified, for example, by first screening for either synthetic or naturally occurring molecules that bind to the recombinantly produced Trp protein and then, in a second step, by testing those selected molecules in cellular assays for inhibition of the Trp protein, as reflected by inhibition of at least one of the biological activities as described in the examples, below. Such screening for molecules that bind Trp8a, Trp8b, Trp9, Trp10a or Trp10b could easily performed on a large scale, e.g. by screening candidate molecules from libraries of synthetic and/or natural molecules. Such an inhibitor is, e.g., a synthetic organic chemical, a natural fermentation product, a substance extracted from a microorganism, plant or animal, or a peptide. Additional examples of inhibitors are specific antibodies, preferably monoclonal antibodies. Moreover, the nucleic sequences of the invention and the encoded proteins can be used to identify further factors involved in tumor development and progression. In this context it should be emphasized that the modulation of the calcium channel of a member of the trp family can result in the stimulation of the immune response of T lymphocytes leading to proliferation of the T lymphocytes. The proteins of the invention can, e.g., be used to identify further (unrelated) proteins which are associated with the tumor using screening methods based on protein/protein interactions, e.g. the two-hybridsystem Fields, S. and Song, O. (1989) Nature (340): 245-246.

The present invention also provides a method for diagnosing a prostate carcinoma which comprises contacting a target sample suspected to contain the protein Trp8a, Trp8b, Trp10a and/or Trp10b or the Trp8a, Trp8b, Trp10a and/or Trp10b encoding mRNA with a reagent which reacts with Trp8a, Trp8b, Trp10a and/or Trp10b or the Trp8a, Trp8b, Trp10a and/or Trp10b encoding mRNA and detecting Trp8a, Trp8b, Trp10a and/or Trp10b or Trp8a, Trp8b, Trp10a and/or Trp10b encoding mRNA.

It has been found that carcinoma cells of placenta (chorion carcinoma), lung and prostate express Trp10 transcripts as well as Trp10 antisense transcripts and transcripts being in part complementary to Trp10 antisense transcripts. Accordingly, the present invention also provides a method for diagnosing a melanoma, chorion carcinoma, cancer of the lung and of the prostate in a tissue of a subject, comprising contacting a sample with a reagent which detects Trp10a and/or Trp10b antisense RNA.

When the target is mRNA (or antisense RNA), the reagent is typically a nucleic acid probe or a primer for PCR. The person skilled in the art is in a position to design suitable nucleic acids probes based on the information as regards the nucleotide sequence of Trp8a, Trp8b, Trp10a or Trp10b as depicted in figure 7, 8a, 10 and 11, respectively, or tables 1 and 2, above. When the target is the protein, the reagent is typically an antibody probe. The term "antibody", preferably, relates to antibodies which consist essentially of pooled monoclonal antibodies with different epitopic specifities, as well as distinct monoclonal antibody preparations. Monoclonal antibodies are made from an antigen containing fragments of the proteins of the invention by methods well known to those skilled in the art (see, e.g., Köhler et al., Nature 256 (1975), 495). As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, Fab and F(ab') 2 fragments) which are capable of specifically binding to protein. Fab and f(ab')2 fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding than an intact antibody. (Wahl et al., J. Nucl. Med. 24: 316-325 (1983)). Thus, these fragments are preferred, as well as the products of a FAB or other immunoglobulin expression library. Moreover, antibodies of the present invention include chimerical, single chain, and humanized antibodies. The target cellular component, i.e. Trp8a, Trp8b, Trp10a and/or Trp10b or Trp8a, Trp8b, Trp10a and/or Trp10b encoding mRNA or Trp10a/b antisense transcripts, e.g., in biological fluids or tissues, may be detected directly in situ, e.g. by in situ hybridization (e.g., according to the examples, below) or it may be isolated from other cell components by common methods known to those skilled in the art before contacting with a probe. Detection methods include Northern blot analysis, RNase protection, in situ methods, e.g. in situ hybridization, in vitro amplification methods (PCR, LCR, QRNA replicase or RNA-transcription/amplification (TAS, 3SR), reverse dot blot disclosed in EP-B1 O 237 362)), immunoassays, Western blot and other detection assays that are known to those skilled in the art.

Products obtained by in vitro amplification can be detected according to established methods, e.g. by separating the products on agarose gels and by subsequent staining with ethidium bromide. Alternatively, the amplified products can be detected by using labeled primers for amplification or labeled dNTPs.

The probes can be detectable labeled, for example, with a radioisotope, a bioluminescent, compound, a chemiluminescent compound, a fluorescent compound, a metal chelate, or an enzyme.

Expression of Trp8a, Trp8b, Trp10a and Trp10b, respectively, in tissues can be studied with classical immunohistological methods (Jalkanen et al., J. Cell. Biol. 101 (1985), 976-985; Jalkanen et al., J. Cell. Biol. 105 (1987), 3087-3096; Sobol et al. Clin. Immunpathol. 24 (1982), 139-144; Sobol et al., Cancer 65 (1985), 2005-2010). Other antibody based methods useful for detecting protein gene expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA). Suitable antibody assay labels are known in the art and include enzyme labels, such as, glucose oxidase, and radioisotopes, such as iodine (125 I, 121 I), carbon (14C), sulfur (35S), tritium (3H), indium (112 In), and technetium rhodamine, and biotin. In addition to assaying Trp8a, Trp8b, Trp 10a or Trp10b levels in a biological sample, the protein can also be detected in vivo by imaging. Antibody labels or markers for in vivo imaging of protein include those detectable by Xradiography, NMR or ESR. For X-radiography, suitable labels include radioisotopes such as barium or cesium, which emit detectable radiation but are not overtly harmful to the subject. Suitable markers for NMR and ESR include those with a detectable characteristic spin, such as deuterium, which may be incorporated into the antibody by labeling of nutrients for the relevant hybridoma. A protein-specific antibody or antibody fragment which has been labeled with an appropriate detectable imaging moiety, such as a radioisotope (for example, <sup>131</sup>L, <sup>112</sup>In, <sup>99</sup>mTc), a radio-opaque substance, or a material detectable by nuclear magnetic resonance, is introduced (for example, parenterally, subcutaneously, or intraperitoneally) into the mammal. It will be understood in the art that the size of the subject and the imaging system used will determine the quantity of imaging moiety needed to produce diagnostic images. In the case of a radioisotope moiety, for a human subject, the quantity of radioactivity injected will normally range from about 5 to 20 millicuries of 99 mTc. The labeled antibody or antibody fragment will then preferentially accumulate at he location of cells which contain the specific protein. In

vivo tumor imaging is described in S.W. Burchiel et al., "Immunopharmacokinetics of Radiolabeled Antibodies and Their Fragments". (Chapter 13 in Tumor Imaging: The Radiochemical Detection of Cancer, S.W. Burchiel and B.A. Rhodes, eds., Masson Publishing Inc. (1982)).

The marker Trp8a and Trp8b is also useful for prognosis, for monitoring the progression of the tumor and the diagnostic evaluation of the degree of malignancy of a prostate tumor (grading and staging), e.g. by using in situ hybridization: In a primary carcinoma Trp8 is expressed in about 2 to 10% of carcinoma cells, in a rezidive carcinoma in about 10 to 60% of cells and in metastases in about 60 to 90% of cells.

The present invention also relates to a method for diagnosing endometrial cancer (cancer of the uterus) which comprises contacting a target sample suspected to contain the protein Trp8a and/or Trp8b or the Trp8a and/or Trp8b encoding mRNA with a reagent which reacts with Trp8a and/or Trp8b or the encoding mRNA and detecting Trp8a and/or Trp8b encoding mRNA. As regards particular embodiments of this method reference is made to the particular embodiments of the method of diagnosing a prostate cancer outlined above.

For evaluating whether the concentration of Trp8a, Trp8b, Trp10a or Trp10b or the concentration of Trp8a, Trp8b, Trp10a or Trp10b encoding mRNA is normal or increased, thus indicative for the presence of a malignant tumor, the measured concentration is compared with the concentration in a normal tissue, preferably by using the ratio of Trp8a:Trp9, Trp8b:Trp9 or Trp10(a or b)/Trp9 for quantification.

Since the prostate carcinoma forms its own basement membrane when growing invasively, it can be concluded that only cells expressing Trp8 and Trp10 are involved in this phenomenon. Thus, it can be concluded that by inhibiting the expression and/or activity of these proteins an effective therapy of cancers like PCA is provided.

Thus, the present invention also relates to a pharmaceutical composition containing a reagent which decreases or inhibits Trp8a, Trp8b, Trp10a and/or Trp10b expression or the activity of Trp8a, Trp8b, Trp10a and/or Trp10b, and a method for preventing, treating, or ameliorating a prostate tumor, endometrial cancer (uterine carcinoma) tumor, a chorion carcinoma, cancer of the lung or melanoma, which comprises administering to a mammalian subject a

**17** 

therapeutically effective amount of a reagent which decreases or inhibits Trp8a, Trp8b, Trp10a and/or Trp10b expression or the activity of Trp8a, Trp8b, Trp10a and/or Trp10b. Examples of such reagents are the above described antisense RNAs, ribozymes or inhibitors, e.g. specific antibodies. Furthermore, peptides, which inhibit or modulate the biological function of Trp8a, Trp8b, Trp9, Trp10a and/or Trp10b may be useful as therapeutical reagents. For example, these peptides can be obtained by screening combina torial phage display libraries (Cosmix, Braunschweig, Germany) as described by Rottgen, P. and Collins, J. (Gene (1995) 164 (2): 243-250). Furthermore, antigenic epitopes of the Trp8 and Trp10 proteins can be identified by the expression of recombinant Trp8 and Trp10 epitope libraries in E. coli (Marquart, A. & Flockerzi, V., FEBS Lett. 407 (1997), 137-140; Trost, C., et al., FEBS Lett. 451 (1999) 257-263 and the consecutive screening of these libraries with serum of patients with cancer of the prostate or of the endometrium. Those Trp8 and Trp10 epitopes which are immunogenic and which lead to the formation of antibodies in the serum of the patients can be then be used as Trp8 or Trp10 derived peptide vaccines for immune inventions against cancer cells which express Trp8 or Trp10. Alternatively to the E. coli expression system, Trp8 or Trp10 or epitopes of Trp8 and Trp10 can be expressed in mammalian cell lines such as human embryonic kidney (Hek 293) cells (American Type Culture Collection, ATCC CRL 1573).

Finally, compounds useful for therapy of the above described diseases comprise compounds which act as antagonists or agonists on the ion channels Trp8, Trp9 and Trp10. It could be shown that Trp8 is a highly calcium selective ion channel which in the presence of monovalent (namely sodium) and divalent ions (namely calcium) is only permeable for calcium ions (see Example 4, below, and Figures 3A, C, E). Under physiological conditions, Trp8 is a calcium selective channel exhibiting large inward currents. This very large conductance of Trp8 channels (as wells as Trp9 and Trp10a/b channels) is useful to establish systems for screening pharmacological compounds interacting with Trp-channels including high throughput screening systems. Useful high throughput screening systems are well known to the person skilled in the art and include, e.g., the use of cell lines stably or transiently transfected with DNA sequences encoding Trp8, Trp9 and Trp10 channels in assays to detect calcium signaling in biological systems. Such systems include assays based on Ca-sensitive dyes such as aequorin, apoaequorin, Fura-2, Fluo-3 and Indo-1.

Accordingly, the present invention also relates to a method for identifying compounds which act as agonists or antagonists on the ion channels Trp8, Trp9 and/or Trp10, said method comprising contacting a test compound with the ion channel Trp8, Trp9 and/or Trp10, preferably by using a system based on cells stably or transiently transfected with DNA sequences encoding Trp8, Trp9 and/or Trp10, and determining whether said test compound affects the calcium uptake.

For administration the above described reagents are preferably combined with suitable pharmaceutical carriers. Examples of suitable pharmaceutical carriers are well known in the art and include phosphate buffered saline solutions, water, emulsions, such as oil/water emulsions, various types of wetting agents, sterile solutions etc. Such carriers can be formulated by conventional methods and can be administered to the subject at a suitable dose. Administration of the suitable compositions may be effected by different ways, e.g. by intravenous, intraperetoneal, subcutaneous, intramuscular, topical or intradermal administration. The route of administration, of course, depends on the nature of the tumor and the kind of compound contained in the pharmaceutical composition. The dosage regimen will be determined by the attending physician and other clinical factors. As is well known in the medical arts, dosages for any one patient depends on many factors, including the patient's size, body surface area, age, sex, the particular compound to be administered, time and route of administration, the kind and stage of the tumor, general health and other drugs being administered concurrently.

The delivery of the antisense RNAs or ribozymes of the invention can be achieved by direct application or, preferably, by using a recombinant expression vector such as a chimeric virus containing these compounds or a colloidal dispersion system. By delivering these nucleic acids to the desired target, the intracellular expression of Trp8a, Trp8b, Trp10a and/or Trp10b and, thus, the level of Trp8a, Trp8b, Trp10a and/or Trp10b can be decreased resulting in the inhibition of the negative effects of Trp8a, Trp8b, Trp10a and/or Trp10b, e.g. as regards the metastasis formation of PCA.

Direct application to the target site can be performed, e.g., by ballistic delivery, as a colloidal dispersion system or by catheter to a site in artery. The colloidal dispersion systems which can be used for delivery of the above nucleic, acids include macromolecule complexes, nanocapsules, microspheres, beads and lipid-based systems including oil-in-water emulsions

(mixed), micelles, liposomes and lipoplexes, The preferred colloidal system is a liposome. The composition of the liposome is usually a combination of phospholipids and steroids, especially cholesterol. The skilled person is in a position to select such liposomes which are suitable for the delivery of the desired nucleic acid molecule. Organ-specific or cell-specific liposomes can be used in order to achieve delivery only to the desired tumor. The targeting of liposomes can be carried out by the person skilled in the art by applying commonly known methods. This targeting includes passive targeting (utilizing the natural tendency of the liposomes to distribute to cells of the RES in organs which contain sinusoidal capillaries) or active targeting (for example by coupling the liposome to a specific ligand, e.g., an antibody, a receptor, sugar, glycolipid, protein etc., by well known methods). In the present invention monoclonal antibodies are preferably used to target liposomes to specific tumors via specific cell-surface ligands.

Preferred recombinant vectors useful for gene therapy are viral vectors, e.g. adenovirus, herpes virus, vaccinia, or, more preferably, an RNA virus such as a Retrovirus. Even more preferably, the retroviral vector is a derivative of a murine or avian retrovirus. Examples of such retroviral vectors which can be used in the present invention are: Moloney murine leukemia virus (MoMuLV), Harvey murine sarcoma virus (HaMuSV), murine mammary tumor virus (MuMTV) and Rous sarcoma virus (RSV). Most preferably, a non-human primate retroviral vector is employed, such as the gibbon ape leukemia virus (GaLV), providing a broader host range compared to murine vectors. Since recombinant retroviruses are defective, assistance is required in order to produce infectious particles. Such assistance can be provided, e.g., by using helper cell lines that contain plasmids encoding all of the structural genes of the retrovirus under the control of regulatory sequences within the LTR. Suitable helper cell lines are well known to those skilled in the art. Said vectors can additionally contain a gene encoding a selectable marker so that the transduced cells can be identified. Moreover, the retroviral vectors can be modified in such a way that they become target specific. This can be achieved, e.g., by inserting a polynucleotide encoding a sugar, a glycolipid, or a protein, preferably an antibody. Those skilled in the art know additional methods for generating target specific vectors. Further suitable vectors and methods for in vitro- or in vivo-gene therapy are described in the literature and are known to the persons skilled in the art; see, e.g., WO 94/29469 or WO 97/00957.

In order to achieve expression only in the target organ, i.e. tumor to be treated, the nucleic acids encoding, e.g. an antisense RNA or ribozyme can also be operably linked to a tissue specific promoter and used for gene therapy. Such promoters are well known to those skilled in the art (see e.g. Zimmermann et al., (1994) Neuron 12, 11-24; Vidal et al.; (1990) EMBO J. 9, 833-840; Mayford et al., (1995), Cell 81, 891-904; Pinkert et al., (1987) Genes & Dev. 1, 268-76).

For use in the diagnostic research discussed above, kits are also provided by the present invention. Such kits are useful for the detection of a target cellular component, which is Trp8a, Trp8b, Trp10a and/or Trp10b or, alternatively, Trp8a, Trp8b, Trp10a and/or Trp10b encoding mRNA or Trp10a/b antisense transcripts, wherein the presence or an increased concentration of Trp8a, Trp8b, Trp10a and/or Trp10b or, alternatively, Trp8a, Trp8b, Trp10a and/or Trp10b encoding mRNA or Trp10a/b antisense transcripts is indicative for a prostate tumor, endometrial cancer, melanoma, chorion carcinoma or cancer of the lung, said kit comprising a probe for detection of Trp8a, Trp8b, Trp9, Trp10a and/or Trp10b or, alternatively, Trp8a, Trp8b, Trp9, Trp10a and/or Trp10b encoding mRNA or Trp10a/b antisense transcripts. The probe can be detectably labeled. Such probe may be a specific antibody or specific oligonucleotide. In a preferred embodiment, said kit contains an anti-Trp8a-, anti-Trp8b-, anti-Trp9-, anti-Trp10a-and/or anti-Trp10b-antibody and allows said diagnosis, e.g., by ELISA and contains the antibody bound to a solid support, for example, a polystyrene microtiter dish or nitrocellulose paper, using techniques known in the art. Alternatively, said kits are based on a RIA and contain said antibody marked with a radioactive isotope. In a preferred embodiment of the kit of the invention the antibody is labeled with enzymes, fluorescent compounds, luminescent compounds, ferromagnetic probes or radioactive compounds. The kit of the invention may comprise one or more containers filled with, for example, one or more probes of the invention. Associated with container (s) of the kit can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, us or sale for human administration.

# **EXAMPLES**

The following Examples are intended to illustrate, but not to limit the invention. While such Examples are typical of those that might be used, other methods known to those skilled in the art may alternatively be utilized.

#### **Example 1: Materials and Methods**

# (A) Isolation of cDNA clones and Northern blot analysis

Total RNA was isolated from human placenta an prostate using standard techniques. Isolation of mRNA was performed with poly (A)<sup>+</sup>RNA - spin columns (New England Biolabs, Beverly, USA) according to the instructions of the manufacturer. Poly (a) <sup>+</sup>RNA was reverse transcribed using the cDNA choice system (Gibco-BRL, Rockville, USA) and subcloned in λ-Zap phages (Stratagene, La Jolla, USA). An human expressed sequence tag (GenBank accession number 1404042) was used to screen an oligo d(T) primed human placenta cDNA library. Several cDNA clones were identified and isolated. Additional cDNA clones were isolated from two specifically primed cDNA libraries using primers 5'-gca tag gaa ggg aca ggt gg-3' and 5'-gag agt cga ggt cag tgg tcc-3'.

cDNA clones were sequenced using a thermocycler (PE Applied Biosystems, USA) and Thermo Sequenase (Amersham Pharmacia Biotech Europe, Freiburg, Germany). DNA sequences were analyzed with an automated sequencer (Licor, Linccoln, USA).

For Northern blot analysis 5  $\mu$ g human poly (A)<sup>+</sup> RNA from human placenta or prostate were separated by electrophoresis on 0.8 % agarose gels. Poly (A)<sup>+</sup> RNA was transferred to Hybond N nylon membranes (Amersham Pharmacia Biotech Europe, Freiburg, Germany). The membranes were hyridized in the presence of 50 % formamide at 42°C over night. DNA probes were labelled using  $[\alpha^{32}P]dCTP$  and the "ready prime,, labelling kit (Amersham Pharmacia Biotech Europe, Freiburg, Germany). Commercial Northern blots were hybridized according to the distributors instructions (Clontech, Paolo Alto, USA).

#### (B) Construction of expression plasmids and transfection of HEK 293 cells

Lipofections were carried out with the recombinant dicistronic eucaryotic expression plasmid pdiTRP8 containing the cDNA of Trp8b under the control of the chicken \( \mathbb{B}\)-actin promotor followed by an internal ribosome entry side (IRES) and the cDNA of the green fluorescent protein (GFP). To obtain pdiTRP8 carrying the entire protein coding regions of TRP8b and

the GFP (Prasher, D.C. et al. (1992), Gene 111, 229-233), the 5'and 3'-untranslated sequences of the TRP8b cDNA were removed, the consensus sequence for initiation of translation in vertebrates (Kozak, M. (1987) Nucleic Acids Research15, 8125-8148) was introduced immediately 5'of the translation initiation codon and the resulting cDNA was subcloned into the pCAGGS vector (Niwa, H., Yamamura, K. and Miyazaki, J (1991), Gene 8, 193-199) downstream of the chicken β-actin promotor. The IRES derived from encephalmyocarditis virus (Kim, D.G., Kang, H.M., Jang, S.K. and Shin H.S. (1992) Mol.Cell.Biol. 12, 3636-3643) followed by the GFP cDNA containing a Ser65Thr mutation (Heim, R., Cubitt, A.B., Tsien, R.Y. (1995) Nature 373, 663-664) was then cloned 3' to the TRP8b cDNA. The IRES sequence allows the simultaneous translation of TRP8b and GFP from one transcript. Thus, transfected cells can be detected unequivocally by the development of green fluorescence.

For monitoring of the intracellular Ca<sup>2+</sup> concentration human embryonic kidney (HEK 293) cells were cotransfected with the pcDNA3-TRP8b vector and the pcDNA3-GFPvector in a molar ratio of 4:1 in the presence of lipofectamine (Quiagen, Hilden, Germany). To obtain pcDNA3-TRP8b the entire protein coding region of TRP8b including the consensus sequence for initiation of translation in vertebrates (Kozak, M. (1987) Nucleic Acids Research15, 8125-8148) was subcloned into the pcDNA3 vector (Invitrogen, Groningen, Netherlands). Calcium monitoring and patch clamp experiments were carried out two days and one day after transfection, respectively.

#### (C) Chromosomal localization of the Trp8 gene

The chromosomal localization of the human TRP8 gene was performed using NIGMS somatic hybrid mapping panel No.2 (Coriell Institute, Camden, NJ, USA) previously described (Drwinga, H.L., Toji, L.H., Kim, C.H., Greene, A.E., Mulivor, R.A. (1993) Genomics 16, 311-314; Dubois, B.L. and Naylor, S.L. (1993) Genomics 16, 315-319).

(D) In Vitro Translation, glutathione - sepharose and calmodulin agarose binding assay N- and C-terminal Trp8-fragments were subcloned into the pGEX-4T2 vector (Amersham Pharmacia Europe, Freiburg, Germany) resulting in glutathione-S-transferase (GST)-Trp8 fusion constructs (Fig. 4). The GST-TRP8-fusion proteins were expressed in E. coli BL 21 cells and purified using glutathione - sepharose beads (Amersham Pharmacia Biotech Europe, Freiburg, Germany).

In vitro translation of human Trp8 cDNA and Xenopus laevis calmodulin cDNA (Davis, T.N. and Thorner, J. Proc.Natl.Acad.Sci. USA 86, 7909-7913.) was performed in the presence of <sup>35</sup>S-methionine using the TNT coupled transcription/translation kit (Promega, Madison, USA). Translation products were purified by gel fliltration (Sephadex G50, Amersham Pharmacia Biotech Europe, Freiburg, Germany) and equal amounts of <sup>35</sup>S labeled probes were incubated for 2 h with glutathione beads bound to GST - Trp8 or calmodulin - agarose (Calbiochem) in 50 mM Tris-HCl, pH 7.4, 0.1 % Triton X-100, 150 mM NaCl in the presence of 1 mM Ca<sup>2+</sup> or 2 mM EGTA. After three washes, bound proteins were eluted with SDS sample buffer, fractionated by SDS-PAGE and <sup>35</sup>S labeled proteins were detected using a Phosphor Imager (Fujifilm, Tokyo, Japan).

# (E) Calcium measurements

The intracellular Ca<sup>2+</sup> concentration ([Ca<sup>2+</sup>]<sub>i</sub>) was determined by dual wavelength fura-2 fluorescence ratio measurements (Tsien, R.Y. (1988) Trends Neurosci. 11, 419-424) using a digital imaging system (T.I.L.L. Photonics, Planegg, Germany). HEK cells were grown in minimal essential medium in the presence of 10 % fetal calf serum and cotransfected with the pcDNA3-TRP8b vector and the pcDNA3-GFPvector as described above (B). Transfected cells were detected by development of green fluorescence. The cells were loaded with 4μM fura-2/AM (Molecular Probes, Oregon, USA) for one hour. After loading the cells were rinsed 3 times with buffer B1 (10 mM Hepes, 115 mM NaCl, 2 mM MgCl<sub>2</sub>, 5mM KCl, pH 7.4) and the [Ca<sup>2+</sup>]<sub>i</sub> was calculated from the fluorescence ratios obtained at 340 and 380 nm excitation wavelengths as described (Garcia, D.E., Cavalié, A. and Lux, H.D. (1994) J. Neurosci 14, 545-553).

#### (F) Electrophysiological recordings

HEK cells were transfected with the eucaryotic expression plasmid pdiTRP8 described in (B) and electrophysiolocigal recordings were carried out one day after transfection. Single cells were voltage clamped in the whole cell mode of the patch clamp technique as described (Hamill, O.P., Marty, A., Neher, E., Sakmann, B. and Sigworth, F.J. (1981) Pflügers Arch. 391, 85-100; Philipp, S., Cavalié, A., Freichel, M., Wissenbach, U., Zimmer, S., Trost, C., Marquart, A., Murakami, M. and Flockerzi, V. (1996) EMBO J. 6166-6171). The pipette solution contained contained (mM): 140 aspartic acid, 10 EGTA, 10 NaCl, 1 MgCl2, 10 Hepes (pH 7.2 with CsOH) or 125 CsCl, 10 EGTA, 4 CaCl<sub>2</sub> 10 Hepes (pH 7,2 with CsOH). The bath solution contained (mM): 100 NaCl, 10 CsCl, 2 MgCl<sub>2</sub>, 50 mannitol, 10 glucose, 20

Hepes (pH 7,4 with CsOH) and 2 CaCl<sub>2</sub>, or no added CaCl<sub>2</sub> (-Ca<sup>2+</sup> solution). Divalent free bath solution contained (mM): 110 N-methyl-D-glucamine (NMDG). Whole cell currents were recorded during 100 msec voltage ramps from -100 to +100 mV at varying holding potentials.

#### (G) In Situ Hybridization

In situ hybridizations were carried out using formalin fixed tissue slices of 6 - 8 µM thickness. The slices were hydrated and incubated in the presence of PBS buffer including 10 µg / ml proteinase K (Roche Diagnostics, Mannheim, Germany) for 0.5 h. The slices were hybridized at 37°C using biotinylated deoxy-oligonucleotides (0.5 pmol / µl) in the presence of 33 % formamide for 12 h. Furthermore the slices were several times rinsed with 2 x SSC and incubated at 25°C for 0.5 h with avidin / biotinylated horse raddish peroxidase complex (ABC, DAKO, Santa Barbara, USA). After several washes with PBS buffer the slices were incubated in the presence of biotinylated tyramid and peroxide (0.15 % w/v) for 10 min, rinsed with PBS buffer and additionally incubated with ABC complex for 0.5 h. The slices were washed with PBS buffer and incubated in the presence of DAB solution (diaminobenzidine (50µg / ml), 50 mM Tris/EDTA buffer pH 8.4, 0.15 % H<sub>2</sub>O<sub>2</sub> in N,N dimethyl-formamide; Merck, Darmstadt, Germany), The detection was stopped after 4 minutes by incubating the slides in water. Tyramid was biotinylated by incubating NHS-LC Biotin (sulfosuccinimidyl-6-(biotinimid)-hexanoat), 2.5 mg/ml; Pierce, Rockford, USA) and tyramin-HCl (0.75 mg / ml, Sigma) in 25 mM borate buffer pH 8.5 for 12 h. The tyramid solution was diluted 1 - 5: 1000 in PBS buffer.

(H) GenBank accession numbers: TRP8a, Aj243500; TRP8b Aj243501

#### **Example 2: Expression of TRP8 transcripts**

In search of proteins distantly related to the TRP family of ion channels, an human expressed sequence tag (EST, GenBank accession number 1404042) was identified in the GenBank database using BLAST programms (at the National Center for Biotechnology Information (NCBI); Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J.J. (1990) Mol. Biol. 5, 403-410) being slightly homologous to the VR1 gene. Several human placenta cDNA libraries were constructed and screeened with this EST DNA as probe. Several full length

cDNA clones were identified and isolated. The full length cDNA clones encoded two putative proteins differing in three amino acids and were termed Trp8a and Trp8b (Fig. 1c, 2a, 7 and 8A). This finding was reproduced by isolating cDNA clones from two cDNA libraries constructed from two individual placentas. The derived protein sequence(s) comprises six transmembrane domains, a characteristic overall feature of trp channels and related proteins (Fig.: 1b). The sequence is closely related to the meanwhile published calcium uptake transport protein 1 (CaT1), isolated from rat intestine (Peng, J.B., Chen, X.Z., Berger, U.V., Vassilev, P.M., Tsukaguchi, H., Brown, E.M. and Hediger M.A. (1999) J Biol Chem. 6;274, 22739-22746) and to the epithelial calcium uptake channel (ECaC) isolated from rabbit kidney (Hoenderop, J.G., van der Kemp, A.W., Hartog, A., van de Graaf, S.F., van Os, C.H., Willems, P.H. and Bindels, R.J. (1999) J Biol Chem. 26;274, 8375-8378). Expression of Trp8a/b transcripts are detectable in human placenta, pancreas and prostate (Fig.: 5) and the size of the Northern signal (3.0 kb) corresponds with the size of the isolated full length cDNAs. In addition, a shorter transcript of 1.8 kb, probably a splice variant, is detectable in human testis. The Trp8 mRNA is not expressed in small intestine or colon (Fig.: 5) implicating that Trp8 is not the human ortholog of the rat CaT1 or rabbit ECaC proteins. To investigate whether there are other related sequences Trp8a/b derived primers (UW241, 5'-TAT GAG GGT TCA GAC TGC-3' and UW242, 5'-CAA AGT AGA TGA GGT TGC-3') were used to amplify a 105 bp fragment from human genomic DNA being 95% identical on the nucleotide level to the Trp8 sequence (data not shown). This indicates the existence of several similar sequences in humans at least at the genomic level.

# Example 3: Two variants of the Trp8 protein (Trp8a and Trp8b) arise by polymorphism

Two variants of the Trp8 cDNA were isolated from human placenta (Fig.: 2A, 7 and 8A) which encoded two proteins which differ in three amino acids and were termed Trp8a and Trp8b. Trp8a/b specific primers were designed to amplify a DNA fragment of 458 bp of the Trp8 gene from genomic DNA isolated from human T-lymphocytes (primer pair: UW243, 5'-CAC CAT GTG CTG CAT CTA CC-3' and UW244, 5'-CAA TGA CAG TCA CCA GCT CC-3'). The amplification product contains a part of the sequence where the derived protein sequence of Trp8a comprises the amino acid valine and the Trp8b sequence methionine as well as a silent base pair exchange (g versus a) and an intron of 303bp (Fig.: 2.A, B). Both variants of the Trp8 genes (a,b) were amplified from genomic DNA in equal amounts indicating the existence of both variants in the human genome and therefore being not the

result of RNA editing (Fig.: 2B). The Trp8a gene can be distinguished from the Trp8b gene by cutting the genomic fragment of 458bp with the restriction enzyme Bsp1286I (Fig. 2B). Using human genomic DNA isolated from blood of twelve human subjects as template, the 458bp fragment was amplified and restricted with BSP1286I. In eleven of the tested subjects only the Trp8b gene is detectable, while one subject (7) contains Trp8a and Trp8b genes (Fig.: 2D). These implicates that the two Trp8 variants arise by polymorphism and do not represent individual genes. Using Trp8 specific primers and chromosomal DNA as template, the Trp8 locus is detectable on chromosome 7 (Fig.: 2C).

# Example 4: Trp8b is a calcium permeable channel

The protein coding sequence of the Trp8b cDNA was subcloned into pcDNA3 vector (Invitrogen, Groningen, Netherlands) under the control of the cytomegalovirus promotor (CMV). Human embryonic kidney (HEK 293) cells were cotransfected with the Trp8b pcDNA3 construct (pcDNA3-Trp8b vector) and the pcDNA3-GFPvector encoding the green fluorescent protein (GFP) in 4:1 ratio. The Trp8b cDNA and the cDNA of the reporter, GFP, was transiently expressed in human embryonic kidney (HEK 293) cells. The intracellular Ca<sup>2+</sup> concentration ([Ca<sup>2+</sup>]<sub>i</sub>) and changes of [Ca<sup>2+</sup>]<sub>i</sub> were determined by dual wavelength fura-2 fluorescence ratio measurements (Fig.: 3F) in cotransfected cells which were identified by the green fluorescence of the reporter gene GFP.

Dual wavelength fura-2 fluorescence ratio measurement is a standard procedure (e.g. in: An introduction of Molecular Neurobiology (ed. Hall, Z.W.)Sinauer Associates, Sunderland, USA (1992)) using fura-2, which is a fluorescent Ca<sup>2+</sup> sensitive dye and which was designed by R.Y.Tsien (e.g. Trends Neurosci. 11, 419-424 (1988) based upon the structure of EGTA. Its fluorescence emission spectrum is altered by binding to Ca<sup>2+</sup> in the physiological concentration range. In the absence of Ca<sup>2+</sup>, fura-2 fluoresces most strongly at an excitation wavelength of 385 nm; when it binds Ca<sup>2+</sup>, the most effective excitation wavelength shifts to 345 nm. This property is used to measure local Ca<sup>2+</sup> concentrations within cells. Cells can be loaded with fura-2 esters (e.g. fura-2AM) that diffuse across cell membranes and are hydrolyzed to active fura-2 by cytosolic esterases.

In the presence of 1mM Ca<sup>2+</sup>, Trp8 expressing cells typically contained more than 300 nM cytosolic Ca<sup>2+</sup>, while non transfected controls contained less than 100 nM Ca<sup>2+</sup> ions (Fig. 3F).

When Trp8b transfected cells were incubated without extracellular Ca<sup>2+</sup>, the intracellular Ca<sup>2+</sup> concentration ([Ca<sup>2+</sup>]<sub>i</sub>) decreased to levels comparable to non transfected cells. Readdition of 1mM Ca<sup>2+</sup> to the bath resulted in significant increase of the cytosolic [Ca<sup>2+</sup>] in Trp8b transfected cells, but not in controls (Fig.: 3F). After readdition of Ca<sup>2+</sup> ions to the bath solution, the cytosolic Ca<sup>2+</sup> concentration remains on a high steady state level in the Trp8b transfected cells.

### Example 5: Trp8 expressing cells show calcium selective inward currents

To characterize in detail the electrophysiological properties of TRP8, TRP8 and GFP were coexpressed in HEK293 cells using the dicistronic expression vector pdiTRP8 and measured currents using the patch clamp technique in the whole cell mode (Hamill, O.P., Marty, A., Neher, E., Sakmann, B. and Sigworth, F.J. (1981) Pflugers Arch., 391, 85-100).

The eucaryotic expression plasmid pdiTRP8 contains the cDNA of Trp8b under the control of the chicken β-actin promotor followed by an internal ribosome entry side (IRES) and the cDNA of the green fluorescent protein (GFP). To obtain pdiTRP8 carrying the entire protein coding regions of TRP8b and the GFP (Prasher, D.C. et al. (1992), Gene 111, 229-233), the 5'and 3'-untranslated sequences of the TRP8b cDNA were removed, the consensus sequence for initiation of translation in vertebrates (Kozak, M. (1987) Nucleic Acids Research15, 8125-8148) was introduced immediately 5'of the translation initiation codon and the resulting cDNA was subcloned into the pCAGGS vector (Niwa, H., Yamamura, K. and Miyazaki, J (1991), Gene 8, 193-199) downstream of the chicken β-actin promotor. The IRES derived from encephalmyocarditis virus (Kim, D.G., Kang, H.M., Jang, S.K. and Shin H.S. (1992) Mol.Cell.Biol. 12, 3636-3643) followed by the GFP cDNA containing a Ser65Thr mutation (Heim, R., Cubitt, A.B., Tsien, R.Y. (1995) Nature 373, 663-664) was then cloned 3' to the TRP8b cDNA. The IRES sequence allows the simultaneous translation of TRP8b and GFP from one transcript. Thus, transfected cells can be detected unequivocally by the development of green fluorescence.

In the presence of 2 mM external calcium, Trp8b transfected HEK cells show inwardly rectifying currents, the size of which depends on the level of intracellular calcium and the electrochemical driving force. The resting membrane potential was held either at -40 mV, or, to lower the driving force for calcium influx in between pulses, at + 70 mV. Current traces

were recorded in response to voltage ramps from -100 to +100 mV, that were applied every second. To monitor inward and outward currents over time, we analyzed the current size at -80 and + 80 mV of the ramps. Figure 3A shows a representative trace of the current at - 80 mV over time. Both at a holding potential of -40 mV or at +70 mV, the currents are significantly larger than in cells transfected with only the GFP containing vector (Fig.: 3E). Interestingly, after changing to a positive holding potential, current size in Trp8 transfected cells slowly increases and reaches steady state after approximately 70 seconds (Fig.: 3A). To determine the selectivity of the induced currents, we then perfused the cells with solutions that either contain no sodium, no added Ca<sup>2+</sup> (Fig. 3A, C) or a sodium containing, but divalent ion free bath solution. To control for the effect of the solution change alone, we also perfused with normal bath (see puff in Fig. 3A). While removal of external Ca<sup>2+</sup> completely abolishes the trp 8 induced currents - the remaining current being identical in size and shape to the control (Fig.: 3A, C, E), removal of external sodium has no effect (Fig.: 3E). An important hallmark of calcium selective channels (e.g. Vennekens, R., Hoenderop, G.J., Prenen, J., Stuiover, M., Willems, PHGM, Droogmans, G., Nilius, B. and Bindels, R.J.M (1999) J. Biol. Chem. 275, 3963-3969), is their ability to conduct sodium only if all external divalent ions, namely Ca<sup>2+</sup> and magnesium are removed. To test whether the trp 8 channel conforms with this phenomenon normal bath solution was switched to a solution containing only sodium and 1 mM EGTA. As can be seen in Figure 3B and D, Trp8 transfected cells can now conduct very large sodium currents. Interestingly, immediately after the solution change, the currents first become smaller before increasing rapidly, indicating that the pore may initially still be blocked by calcium a phenomenon usually called anomalous mole fraction behaviour (Warnat, J., Philipp, S., Zimmer, S., Flockerzi, V., and Cavalié A.(1999) J.Physiol. (Lond) 518, 631-638). The measured outward currents of Trp8 transfected cells in normal bath solution are not significantly different from non-transfected control cells or cells which only express the reporter gene GFP. As the removal of external Ca2+ abolishes the Trp8 specific current, the remaining current was subtracted from the current before the solution change to obtain the uncontaminated Trp8 conductance (see inset in Fig.: 3C). As expected from the given ionic conditions (high EGTA inside, 2 mM Ca<sup>2+</sup> outside), the current-voltage relationship now shows prominent inward rectification with little to no outward current.

Both the time course of the development of Trp8 currents and the size of the currents depend on the frequency of stimulation (data not shown), the internal and external Ca<sup>2+</sup> concentration

and the resting membrane potential, suggesting that Trp8 calcium conductance is intrically regulated by a Ca<sup>2+</sup> mediated feedback mechanisms.

# Example 6: Ca<sup>2+</sup> / calmodulin binds to the C-terminus of the Trp8 protein

To test whether calmodulin, a prime mediator of calcium regulated feedback, is involved, first it was investigated biochemically whether Trp8 protein can bind calmodulin. Trp8 cDNA was in vitro translated in the presence of <sup>35</sup>S-methionine and the product incubated with calmodulin-agarose beads. After several washes either in the presence or abscence of Ca<sup>2+</sup>, the beads were incubated in Laemmli buffer and subjected to SDS-polyacrylamide gel electrophoresis. In the presence of Ca<sup>2+</sup> (1mM), but not in the absence of Ca<sup>2+</sup>, Trp8 protein binds to calmodulin (Fig.: 4B).

To narrow down the binding site, two approaches were undertaken: Firstly, GST-TRP8 fusion proteins of various intracellular domains of Trp8 were constructed, expressed in E. coli and bound to gluthathione sepharose beads. These beads were then incubated with in vitro translated <sup>35</sup>S- labeled calmodulin, washed and subjected to gel electrophoresis. Secondly, truncated versions of in vitro translated Trp8 protein were used in the above described binding to calmodulin-agarose. As shown in Figure 4A, and C, fusion proteins of the N-terminal region (N1, N2) of Trp8 did not bind calmodulin, while C-terminal fragments (C1, C2, C3, C4) showed calmodulin binding in the presence of calcium (for localization of fragments within the entire Trp8 protein see Fig. 4C). Accordingly, a truncated version of in vitro translated Trp8, which lacks the C-terminal 32 amino acid residues did not bind to calmodulin-agarose (4B). We have restricted the calmodulin binding site to amino acid residues 691 to 711 of the Trp8 protein. This calmodulin binding site does not resemble the typical conserved IQ - motif of conventional myosins, but has limited sequence homology to the calcium dependent calmodulin binding site 1 of the transient receptor potential like (trpl) protein of Drosophila melanogaster (Warr and Kelly, 1996) with several charged amino acid residues conserved. The sequence of the calmodulin binding site of the Trp8 protein resembles a putative amphipathic α-helical wheel structure with a charged and a hydrophobic site according to a model proposed by Erickson-Vitanen and De Grado (1987, Methods Enzymol. 139, 455-478.).

# Example 7: Expression of Trp8 transcripts in human placenta and pancreas

Several slides from a human placenta of a ten week old abort were used for in situ hybridization experiments. The in situ hybridization experiments revealed expression of Trp8 transcripts in human placenta (Fig.: 5B). Expression was detectable in trophoblasts and syncytiotrophoblasts of the placenta, but not in Langhans cells.

Trp8 transcripts are detectable in human pancreas (Fig.: 5A). Therefore Trp8 probes were hybridized to tissue sections of human pancreas. The pancreatic tissues were removed from patients with pancreas cancer. Trp8 expression is detectable in pancreatic acinar cells, but not in Langerhans islets (Fig.: 5C). No Trp8 expression was found in regions of pancreatic carcinomas (data not shown).

Furthermore, the Trp8 cDNA is not detectable in human colon nor in human kidney by in situ hybridization as well as by Northern analysis (Fig.: 5A, D). The Northern results taken together with the in situ expression data indicate that the Trp8 protein is not the human ortholog of the CaT1 and ECaC channels cloned from rat intestine (Peng, J.B., Chen, X.Z., Berger, U.V., Vassilev, P.M., Tsukaguchi, H., Brown, E.M. and Hediger M.A.(1999) J Biol Chem. 6;274, 22739-22746) and from rabbit kidney (Hoenderop, J.G., van der Kemp, A.W., Hartog, A., van de Graaf, S.F., van Os, C.H., Willems, P.H. and Bindels, R.J. (1999) J Biol Chem. 26;274, 8375-8378), respectively. Trp8 is unlikely to represent the human version of CaT1 as its expression is undetectable in the small intestine and colon tissues where CaT1 is abundantly expressed. If, however, Trp8 is the human version of rat CaT1, a second gene product appears to be required for Ca<sup>2+</sup> uptake in human small intestine and colon attributed to CaT1 in rat small intestine and colon.

# Example 8: Differential expression of Trp8 transcripts in benign and malign tissue of the prostate

The Trp8 transcripts are expressed in human prostate as shown by hybridization of a Trp8 probe to a commercial Northern blot (Clontech, Palo Alto, USA) (Fig.: 5A). Trp8 transcripts were not detectable by Northern blot analysis using pooled mRNA of patients with benign prostatic hyperplasia (BPH) (Fig.: 5A, prostate\*). To examine Trp8 expression on the cellular

level, sections of prostate tissues were hybridized using Trp8 specific cDNA probes (Table 3). Expression of Trp8 transcripts is not detectable in normal prostate (n = 3), benign hyperplasia (BPH, n = 15) or prostatic intraepithelial neoplasia (PIN, n = 9) (Fig.: 6A, C, E). Trp8 transcripts were only detectable in prostate carcinoma (PCA), although with different expression levels. Low expression levels were found in primary carcinomas (2 - 10 % of the carcinoma cells, n = 8) (Fig.: 7B). Much stronger expression was detectable in rezidive carcinoma (10 - 60 %) (Fig.: 7D, n = 6) and metastases of the prostate (60 - 90 %, n = 4) (Fig.: 7F). Thus it has to be concluded that the commercial Northern blot used in Fig.: 5A contains not only normal prostate mRNA as indicated by the distributor. According to the distributors instructions the prostate mRNA used for this Northern blot was collected from 15 human subjects in the range of 14 to 60 years of age. This prostate tissue was not examined by pathologic means. Since Trp8 expression is not detectable in normal or benign prostate, this finding implicates that the mRNA used for this Northern blot was extracted in part from prostatic carcinoma tissue. To summarize, Trp8 expression is only detectable in malign prostate and, thus, the Trp8 cDNA is a marker for prostate carcinoma. The results are summarized in Table 4.

Table 3

Trp8 probes used for in situ hybridization:

Probes (antisense)

- 1.) 5' TCCGCTGCCGGTTGAGATCTTGCC 3'
- 2.) 5' CTTGCTCCATAGGCAGAGAATTAG 3'
- 3.) 5' ATCCTCAGAGCCCCGGGTGTGGAA3'

#### Controls (sense)

- 1.) 5' GGCAAGATCTCAACCGGCAGCGGA 3'
- 2.) 5' CTAATTCTCTGCCTATGGAGCAAG 3'
- 3.) 5' TTCCACACCCGGGGCTCTGAGGAT 3'

Table 4

| Prostate | total | negative | positive |
|----------|-------|----------|----------|
| normal   | 3     | 3        | 0        |
| BPH      | 15    | 15       | 0        |
| PIN      | 9     | 9        | 0        |

carcinoma 18 1 17

# (B) Differential expression of Trp8 transcripts in benign and malign tissue of the uterus

Moreover it could be shown that Trp8 is expressed in endometrial cancer (also called cancer of the uterus, to be distinguished from uterine sarcoma or cancer of the cervix) whereas no expression was observed in normal uterus tissue. Thus, Trp8 also is a specific marker for the diagnosis of the above cancer (Fig. 12).

# **Example 9: Characterization of Trp9**

The complete protein coding sequence of Trp9 was determined (Fig. 9). Trp 9 transcripts are predominantly expressed in the human prostate and in human colon. As it could be shown by Northern blot analysis, there is no difference of the expression of TRP9 in benigne prostata hyperplasia (BPH, Fig. 13, upper panel left) or prostate carcinoma (Fig. 13, upper panel right). However, Trp9 is useful as a reference marker for prostata carcinoma, i.e. can be used for quantifying the expression level of Trp8. The ratio of the expression of Trp8:Trp9 in patients and healthy individuals is useful for the development of a quantitative assay.

# Example 10: Characterization of Trp10

The complete protein coding sequence of TRP10 (a and b) was determined by biocomputing (Fig. 10 and 11). Using a 235 bp fragment of the Trp10 cDNA as probe in Northern blot analysis TRP10 transcripts could only be detected in mRNA isolated from individuals with prostate cancer (Fig. 13, bottom panel) but not in mRNA isolated from benign tissue of the prostate (prostate BPH) nor in mRNA isolated from heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas. The 235 bp cDNA fragment of the Trp10 cDNA was amplified using the primer pair UW248 5'-ACA GCT GCT GGT CTA TTC C-3' and UW249 5'-TAT

GTG CCT TGG TTT GTA CC-3' and prostate cDNA as template. In summary, Trp10a and Trp10b, like TRP8 are also expressed in malignant prostate tissue. So far, its expression could not be observed in any other tissue examined (see above). Thus, Trp 10a and Trp10b are also useful markers which are specific for malignant prostate tissue.

Furthermore, database searches in public databases of the national center for biological information (NCBI) revealed the existence of several expressed sequence tags (EST clones) being in part identical to the Trp10 sequence. These EST clones were originally isolated from cancer tissues of lung, placenta, prostate and from melanoma. These clones include the clones with the following accession numbers: BE274448, BE408880, BE207083, BE791173, AI671853, BE390627. The results demonstrate that cancer cells of these tissues express Trp10 related transcripts whereas no expression of Trp10 transcripts in the corresponding healthy tissues are detectable (Figure 13). Furthermore, it could be shown that in cancer cells of melanoma and prostate cancer Trp10 transcripts are expressed as shown by in situ hybridizations using 4 antisense probes (Figure 14A - E and 13K-O and Table 2, above). Furthermore, it could clearly be shown that cancer cells of these tissues expressing Trp10 transcripts also express Trp10-antisense transcripts as shown in Figure 14F-J, Figure 14P-R and Figure 14T by in situ hybridizations using 4 sense probes (Table 2, above). The in situ hybridization experiments demonstrate that detection of a subset of cancer cells derived from carcinoma of lung, placenta, prostate and melanoma is feasible using antisense as well as sense probes complementary to Trp10 transcripts or complementary to Trp10-antisense transcripts, respectively.

The foregoing is meant to illustrate but not to limit the scope of the invention. The person skilled in the art can readily envision and produce further embodiment, based on the above teachings, without undue experimentation.

#### What Is claimed Is:

1. An isolated nucleic acid molecule encoding the human prostate carcinoma associated protein Trp8a, Trp8b, Trp9, Trp10a or Trp10b or a protein exhibiting biological properties of Trp8a, Trp8b, Trp9, Trp10a or Trp10b and being selected from the group consisting of

- (a) a nucleic acid molecule encoding a protein that comprises the amino acid sequence depicted in Figure 7, 8A, 9, 10 or 11;
- (b) a nucleic acid molecule comprising the nucleotide sequence depicted in Figure 7, 8A, 9, 10 or 11;
- (c) a nucleic acid molecule included in DSMZ Deposit No. DSM 13579, DSM 13580, DSM 13584, DSM 13581 or DSM....;
- (d) a nucleic acid molecule which hybridizes to a nucleic acid molecule specified in (a) to (c);
- (e) a nucleic acid molecule the nucleic acid sequence of which deviates from the nucleic sequences specified in (a) to (d) due to the degeneration of the genetic code; and
- (f) a nucleic acid molecule, which represents a fragment, derivative or allelic variation of a nucleic acid sequence specified in (a) to (e).
- 2. A recombinant vector containing the nucleic acid molecule of claim 1
- 3. The recombinant vector of claim 2 wherein the nucleic acid molecule is operatively linked to regulatory elements allowing transcription and synthesis of a translatable RNA in prokaryotic and/or eukaryotic host cells.
- 4. A recombinant host cell which contains the recombinant vector of claim 3.
- 5. The recombinant host cell of claim 4, which is a mammalian cell, a bacterial cell, an insect cell or a yeast cell.
- 6. An isolated protein exhibiting biological properties of the human prostate carcinoma associated protein Trp8a, Trp8b, Trp9, Trp10a or Trp10b which is encoded by a nucleic acid molecule of claim 1.
- 7. A recombinant host cell that expresses the isolated protein of claim 6.

8. A method of making an isolated protein exhibiting biological properties of the human prostate carcinoma associated protein Trp8a, Trp8b, Trp9, Trp10a or Trp10b comprising: (a) culturing the recombinant host cell of claim 6 under conditions such that said protein is expressed; and

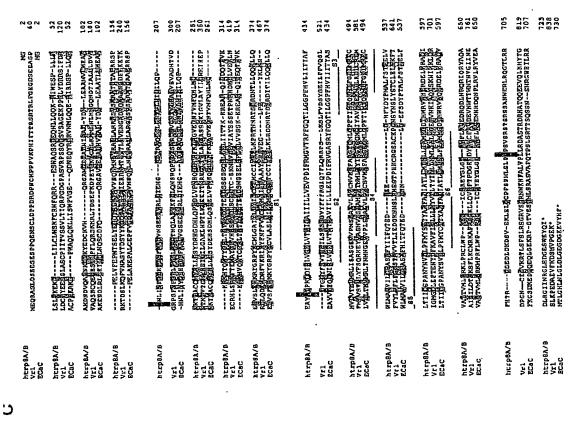
- (b) recovering said protein.
- 9. The protein produced by the method of claim 8.
- 10. An antisense RNA sequence characterized in that it is complementary to an mRNA transcribed from a nucleic acid molecule of claim 1 or a part thereof and can selectively bind to said mRNA or part thereof, said sequence being capable of inhibiting the synthesis of the protein encoded by said nucleic acid molecule.
- 11. A ribozyme characterized in that it is complementary to an mRNA transcribed from a nucleic acid molecule of claim 1 or a part thereof and can selectively bind to and cleave said mRNA or part thereof, thus inhibiting the synthesis of the protein encoded by said nucleic acid molecule.
- 12. An inhibitor characterized in that it can suppress the activity of the protein of claim 6.
- 13. A method for diagnosing a prostate carcinoma which comprises contacting a target sample suspected to contain the protein Trp8a, Trp8b, Trp10a and/or Trp10b or the Trp8a, Trp8b, Trp10a and/or Trp10b encoding mRNA with a reagent which reacts with Trp8a, Trp8b, Trp10a and/or Trp10b or the Trp8a, Trp8b, Trp10a and/or Trp10b encoding mRNA and detecting Trp8a, Trp8b, Trp10a and/or Trp10b or the Trp8a, Trp8b, Trp10a and/or Trp10b encoding mRNA.
- 14. The method of claim 13, wherein the reagent is a nucleic acid.
- 15. The method of claim 13, wherein the reagent is an antibody.
- 16. The method of claim 13, wherein the reagent is detectably labeled.

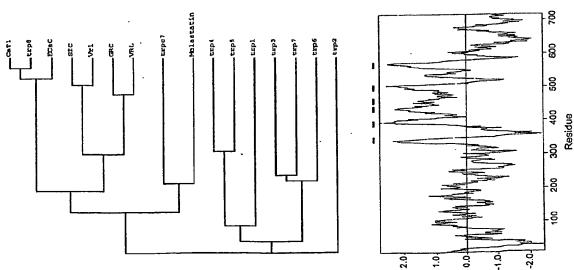
17. The method of claim 16, wherein the label is selected from the group consisting of a radioisotope, a bioluminescent compound, a chemiluminescent compound, a fluorescent compound, a metal chelate, or an enzyme.

- 18. A method for diagnosing an endometrial cancer (carcinoma of the uterus) which comprises contacting a target sample suspected to contain the protein Trp8a and/or Trp8b or the Trp8a and/or Trp8b encoding mRNA with a reagent which reacts with Trp8a and/or Trp8b or the Trp8a and/or Trp8a and/or trp8b encoding mRNA and detecting Trp8a and/or Trp8b or the Trp8a and/or Trp8b encoding mRNA.
- 19. The method of claim 18, wherein the reagent is a nucleic acid.
- 20. The method of claim 18, wherein the reagent is an antibody.
- 21. The method of claim 18, wherein the reagent is detectably labeled.
- 22. The method of claim 21, wherein the label is selected from the group consisting of a radioisotope, a bioluminescent compound, a chemiluminescent compound, a fluorescent compound, a metal chelate, or an enzyme.
- 23. A method for diagnosing a melanoma, chorion carcinoma, cancer of the lung and of the prostate in a tissue of a subject, comprising contacting a sample with a reagent which detects Trp10a and/or Trp10b antisense RNA or Trp10a and/or Trp10b related antisense RNA.
- 24. A method for preventing, treating, or ameliorating a prostate tumor, endometrial cancer (carcinoma of the uterus) tumor, a chorion carcinoma, cancer of the lung or melanoma, which comprises administering to a mammalian subject a therapeutically effective amount of a reagent which decreases or inhibits expression of Trp8a, Trp8b, Trp10a and/or Trp10b and/or the activity of Trp8a, Trp8b, Trp10a and/or Trp10b.
- 25. The method of claim 24, wherein the reagent is a nucleotide sequence comprising an antisense RNA.

26. The method of claim 24, wherein the reagent is a nucleotide sequence comprising a ribozyme.

- 27. The method of claim 24, wherein the reagent is an inhibitor of Trp8a, Trp8b, Trp10a and/or Trp10b.
- 28. The method of claim 27, wherein the reagent is an anti-Trp8a-, anti-Trp10a-and/or anti-Trp10b antibody or a fragment thereof.
- 29. A diagnostic kit useful for the detection of Trp8a, Trp8b, Trp10a and/or Trp10b or Trp8a, Trp8b, Trp10a and/or Trp10b encoding mRNA or Trp10a and/or Trp10b antisense transcripts in a sample, wherein the presence of an increased concentration of Trp8a, Trp8b, Trp9, Trp10a and/or Trp10b or Trp8a, Trp8b, Trp9, Trp10a and/or Trp10b encoding mRNA or Trp10a and/or Trp10b antisense transcripts is indicative for a prostate tumor, endometrial cancer (cancer of the uterus) tumor, a chorion carcinoma, cancer of the lung or melanoma, said kit comprising a probe for detection of Trp8a, Trp8b, Trp9, Trp10a or Trp10b or Trp8a, Trp8b, Trp9, Trp10a and/or Trp10b encoding mRNA or Trp10a and/or Trp10b antisense transcripts.
- 30. The kit of claim 29, wherein the target component to be detected is Trp8a, Trp8b, Trp9, Trp10a and/or Trp10b and the probe is an antibody.
- 31. A method for identifying a compound which acts as an agonist or antagonist on the ion channels Trp8, Trp9 and/or Trp10, said method comprising contacting a test compound with the ion channel Trp8, Trp9 and/or Trp10, and determining whether said test compound affects the calcium uptake.





σ.

ig. 1

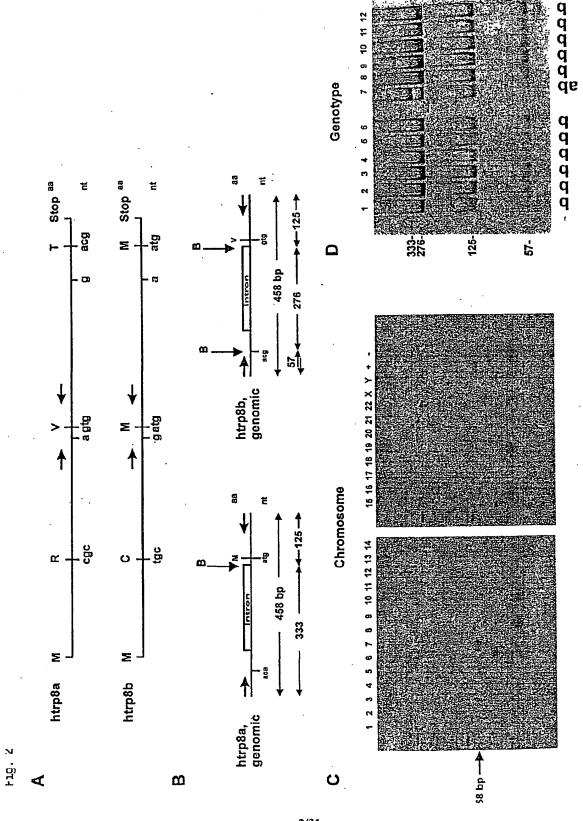


Fig. 3

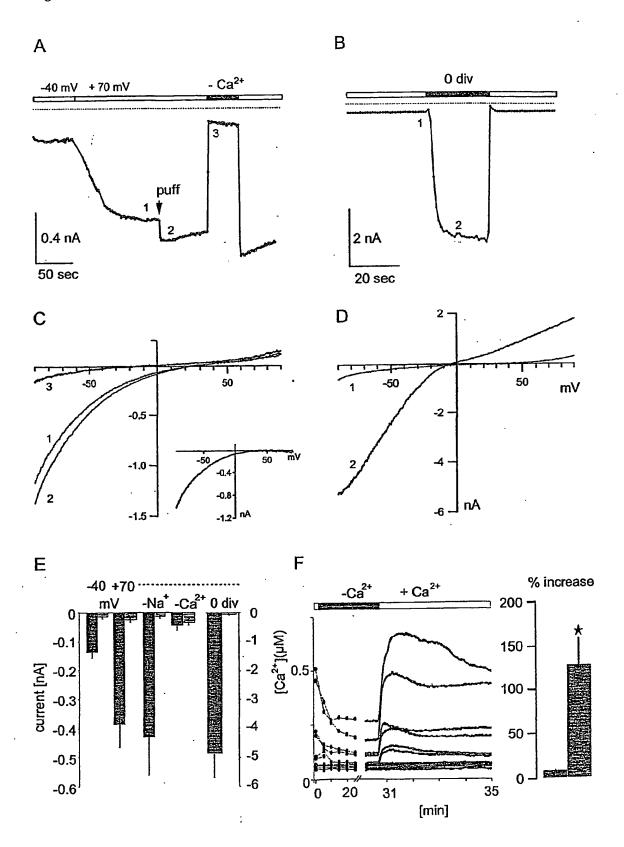
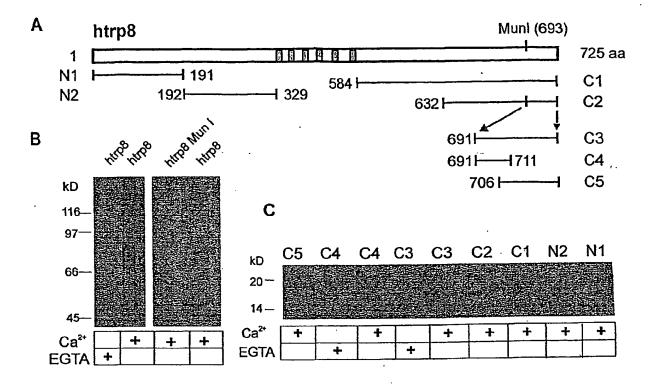


Fig. 4



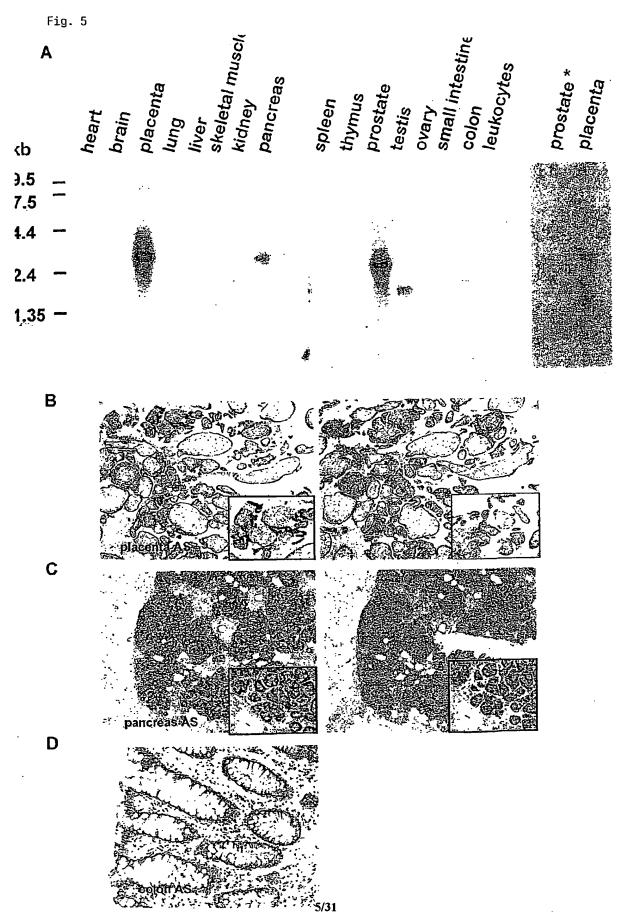


Fig. 6

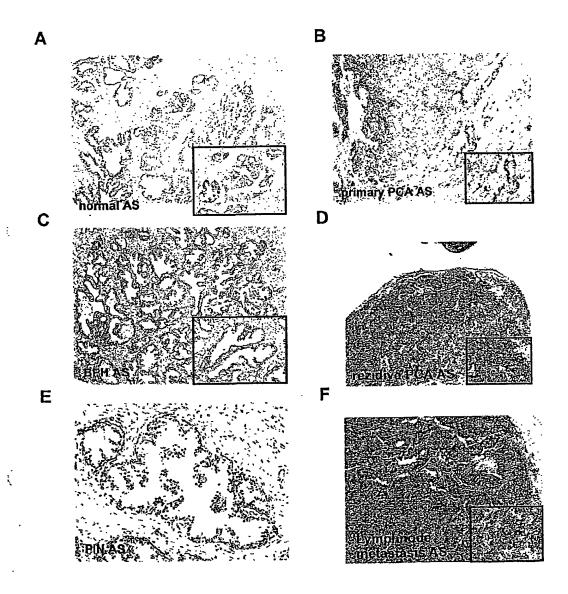


Fig. 7

|   |  | 10  |  |  |  |  |   | 30  |  |  |  |  |   | 5  | 50   |  |  |  |
|---|--|---|--|--|--|--|---|---|--|--|--|--|---|--|--|--|--|--|
| GCCZ  | AAGI   | rgtaaca   | AAAC   | TCA  | CAG  | ccc  | стс   |   | AAC  | TGO  | CTC  | GGG  | CTO                                       | _  |  | 4GA(   | TCC  | CA   |
| 000.  |  | 70  |  |  |  |  |   | 90  |  | -  |  |  |   | 1.3  |  |  |  |  |
| AGG   | AAC:   | CGTCA   | GAA  | GGC  | AGG  | AGI  | ACA   | GGAG  | AΩ   | GG2  | ACCI   | rct <i>i</i>   | ACAC                                      | GGZ  | AGA(   | CGGT   | GGG  | SCC  |
|   |  | 130   |  |  |  |  |   | 150   |  |  |  |  |   |  | 70   |  |  |  |
| GGC   | CCT  | regege  | GCT  | GAT  | GTO  | GCC  | ccc   | AAGG  | CTC  | AG:  | rcc  | CGT  | CAG                                       | GT   | CTG  | GCC?   | rcg(   | 3CC  |
|   |  | 190   |  |  |  |  |   | 210   |  |  |  |  |   | 23   | 30   |  |  |  |
| TCA   | GGC  | CCCAA   | GGAG   | CCG  | GC   | CTZ  | ACA   | cccc  | ATC  | GG:  | rtt(   | GTC2   | ACT(                                      | 3CC(   | CAA  | GGA(   | SAAZ   | AGG  |
|   |  |   |  |  |  |  |   |   | M  | G  | L  | s  | L   | P  | K  | E  | K  | G  |
|   |  | 250   |  |  |  |  |   | 270   |  |  |  |  |   | 2  | 90   |  |  |  |
| GCT   | TAA  | TCTCTG  | CCTA   | \TGG   | AG   | CAAC   | GTT   | CTGC  | AG   | ATG(   | GTT(   | CCA  | GAG                                       | ACG(   | GGA(   | GTC  | CTG  | GGC  |
| L   | I  | L C   | L  | W  | S  | K  | F   | С   | R  | W  | F  | Q  | R   | R  | E  | S  | W  | A  |
|   |  | 310   |  |  |  |  |   | 330   |  |  |  |  |   | 3  | 50   |  |  |  |
| CCA   | GAG  | CCGAGA'   | TGAG   | CAG  | AA   | CCT  | GCI   | GCAG  | CA   | GAA  | GAG  | GAT  | CTG                                       | GGA  | GTC  | TCC  | CTC  | CCT  |
| Q   | S  | R D   | E  | Q  | N  | L  | L   | Q   | Q  | K  | R  | I  | W   | E  | S  | P  | L  | L  |
|   |  | 370   |  |  |  |  |   | 390   |  |  |  |  |   | -  | 10   |  |  |  |
| TCT   | AGC  | TGCCAA  | AGAI   | raat   | 'GA'   | TGT  | CCP   | 4GGCC   | CT   | GAA  | CAA  | GTT  | GCT                                       | CAA  | GTA  | TGA  | GGA'   | TTG  |
| L   | A  | A K   | D  | Ν.   | D  | V  | Q   | A   | F  | N  | K  | L  | L   | K  | Y  | E  | D.   | С  |
|   |  | 430   |  |  |  |  |   | 450   |  | ٠  |  |  |   | -  | 70   |  |  |  |
| CAA   | GGT  | GCACCA  | GAG  | AGG?   | /GC  | CAT  | GGG   | GGAZ  | AAC.   | AGC  | GCT  | ACA  | CAT                                       | AGC.   |  |  |  |  |
| K   | V  | H Q   | R  | G  | A  | M  | G   |   | T  | A  | L  | H  | I   | A  | A  | L  | Y  | D  |
|   |  | 490   |  |  |  |  |   | 510   |  |  |  |  |   | _  | 30   |  |  |  |
| CAA   | CCT  | GGAGGC  | CGC(   |  |  |  |   |   |  |  |  |  |   |  |  |  |  |  |
| N   | L  | E A   | A  | М  | V  | L  | M   | E   | A  | A  | Р  | E  | L   | V_   | _  | E  | P  | M  |
|   |  | 550   |  |  |  |  |   | 570   |  |  |  |  |   | -  | 90   |  |  |  |
|   |  | TGAGCT  |  |  |  |  |   |   |  |  |  |  |   |  |  |  |  |  |
| T   | S  | E L   | Y  | E  | G  | Q  | T   | A   | L  | H  | I  | A  | V   | ٧  |  | Q  | N  | M  |
|   |  | 610   |  |  |  |  |   | 630   |  |  |  |  |   | _  | 50   |  |  | ~~~  |
|   |  | GGTGCG  |  |  |  |  |   |   |  |  |  |  |   |  |  | T  | AGG<br>G   | T  |
| N   | L  | V R   | A  | L  | L  | A  | R   | R<br>690  | A  | S  | V  | S  | A   | R  | A<br>10  | T  | G  | 1  |
|   |  | 670   |  |  |  |  |   |   |  |  |  |  |   |  |  |  |  |  |
|   |  |   | m 3 00   | m  |  | ~~~  | -   |   | <b>~</b> ⊞ 7.  | con  | w.c.c  | ירייזי   | con                                       | -  |  | YCTP (*  | - Trum   | יזיכר  |
|   |  | CCCCC   |  |  |  |  |   | CAT   |  |  |  |  |   | CCC  | TTT  |  |  |  |
| TGC<br>A  | CTI<br>F   | R R   | TAG:   | TCC(   | CCG<br>R   | CAA<br>N   | CC!   | ICAT  | CTA<br>Y   | CTI<br>F   | TGG<br>G   | GGA<br>E   | GCA<br>H                                  | P  | TTI<br>L   | GTC<br>S   | CTT<br>F   | TGC<br>A   |
| A   | F  | R R<br>730  | S  | P  | R  | N  | L   | TCAT<br>I<br>750  | Y  | F  | G  | E  | Н   | CCC<br>P   | TTI<br>L   | s  | F  | A  |
| A<br>TGC  | F<br>CTC   | R R<br>730<br>TGTGAP  | s<br>Cag   | P<br>TGA   | R<br>GGA   | n<br>gat   | L<br>CG   | ICAT<br>I<br>750<br>IGCG  | Y<br>GCT   | e<br>Ect   | G<br>CAT   | E  | Н   | CCC<br>P   | TTI<br>L   | s  | F  | A  |
| A   | F  | R R<br>730<br>STGTGAA<br>V N  | S  | P  | R  | N  | L   | ICATO<br>I<br>750<br>IGCG<br>R  | Y<br>GCT<br>L  | F  | G  | E<br>TGA   | H<br>GCA                                  | P<br>7<br>TGG  | TTI<br>L<br>70<br>SAGO   | S<br>TGA   | F<br>CAT   | A<br>CCG   |
| TGC<br>A  | F<br>CTC   | R R<br>730<br>TGTGAP<br>V N<br>790  | s<br>ACAG<br>S   | P<br>TGA<br>E  | R<br>GGA<br>E  | N<br>GAT<br>I  | L<br>CG<br>V  | ICATO<br>750<br>IGCG<br>R<br>810  | Y<br>GCT<br>L  | F<br>GCI<br>L  | G<br>CAT<br>I  | E<br>TGA<br>E  | H<br>LGCA<br>H                            | P<br>7<br>TGG<br>G   | TTT<br>L<br>70<br>GAGC<br>A  | S<br>TGA<br>D  | F<br>CAT<br>I  | A<br>CCCG<br>R   |
| TGC<br>A  | F<br>C<br>C  | R R<br>730<br>TGTGAP<br>V N<br>790<br>AGGACTO   | s<br>ACAG<br>S   | P<br>TGA<br>E  | R<br>GGA<br>E  | N<br>GAT<br>I  | L<br>CG<br>V  | ICATO<br>TOO<br>TGCG<br>R<br>810  | Y<br>GCT<br>L  | F<br>GCI<br>L  | G<br>CAT<br>I  | E<br>TGA<br>E  | H<br>LGCA<br>H                            | P<br>7<br>TGG<br>G   | TTTI<br>L<br>70<br>AGC   | S<br>TGA<br>D  | F<br>CAT<br>I  | A<br>CCCG<br>R   |
| TGC<br>A  | F<br>CTC   | R R<br>730<br>TGTGAP<br>V N<br>790  | S<br>CAG<br>S  | P<br>TGAG<br>E<br>GGG                                      | R<br>GGA<br>E<br>AAA   | N<br>GAT<br>I<br>CAC   | L<br>CG<br>V  | ICATO<br>TOO<br>TGCG<br>R<br>810  | Y<br>GCT<br>L<br>ACA<br>H  | F<br>GCI<br>L<br>CAI   | G<br>'CA'I<br>I  | E<br>TGA<br>E  | H<br>GCA<br>H                             | TGG  | TTTI<br>L<br>70<br>AGC   | S<br>TGA<br>D<br>CCAA                                | F<br>CAT<br>I  | A<br>CCCG<br>R<br>LAAC   |
| TGC A GGC   | F<br>C<br>C<br>C<br>C<br>C<br>C                          | R R 730 TGTGAP V N 790 AGGACTO D S 850  | S<br>ACAG<br>S<br>S<br>CCT<br>L  | P<br>TGA<br>E<br>GGG                                       | R<br>GGA<br>E<br>AAA<br>N  | N<br>GAT<br>I<br>CAC   | L<br>CG:<br>V<br>AG:<br>V   | ICATO I 750 IGCG R 810 IGTT L 870   | Y<br>GCT<br>L<br>ACA<br>H  | GCT<br>L<br>CAT  | G<br>CAT<br>I<br>CCT   | E<br>TGA<br>E<br>CAI   | H<br>H<br>CCT                             | TGG  | TTT<br>L<br>70<br>SAGO<br>A<br>30<br>AGO<br>P  | S<br>ETGA<br>D<br>CCAA                               | F<br>CAT<br>I<br>ACAA<br>K   | A<br>CCCG<br>R<br>NAAC<br>T  |
| TGC A GGC   | F<br>C<br>C<br>C<br>C<br>C<br>C                          | R R<br>730<br>TGTGAP<br>V N<br>790<br>AGGACTO<br>D S  | S<br>ACAG<br>S<br>S<br>CCT<br>L  | P<br>TGA<br>E<br>GGG                                       | R<br>GGA<br>E<br>AAA<br>N  | N<br>GAT<br>I<br>CAC   | L<br>CG:<br>V<br>AG:<br>V   | ICATO I 750 IGCG R 810 IGTT L 870 TGCT  | Y<br>GCT<br>L<br>ACA<br>H  | GCT<br>L<br>CAT  | G<br>CAT<br>I<br>CCT   | E<br>TGA<br>E<br>CAI   | H<br>H<br>CCT                             | TGG  | TTT<br>L<br>70<br>SAGO<br>A<br>30<br>AGO<br>P  | S<br>ETGA<br>D<br>CCAA                               | F<br>CAT<br>I<br>ACAA<br>K   | A<br>CCCG<br>R<br>NAAC<br>T  |
| TGC A GGC A   | F<br>C<br>C<br>C<br>C<br>C<br>C<br>Q                     | R R 730 TGTGAP V N 790 AGGACTO D S 850  | S<br>ACAG<br>S<br>CCT<br>L   | P<br>TGAG<br>E<br>GGG<br>G                                 | R<br>GGA<br>E<br>AAA<br>N  | GAT I CAC T  | L<br>CG:<br>V<br>AG:<br>V   | ICATO I 750 IGCG R 810 IGTT L 870 TGCT  | Y<br>GCT<br>L<br>ACA<br>H<br>GTC                                     | GCI<br>L<br>CAI<br>I   | G<br>CAT<br>I<br>CCT<br>L  | E<br>TGA<br>E<br>CAT<br>I  | H<br>GCA<br>H<br>CCI<br>L                 | TCCA   | TTT<br>L<br>70<br>GAGC<br>A<br>130<br>AGCC<br>P<br>190<br>GGG  | S<br>CTGA<br>D<br>CCAA<br>N                          | F<br>CAT<br>I<br>CAA<br>K  | A<br>CCCG<br>R<br>LAAC<br>T  |
| A TGC A GGC A CTT F                                   | F<br>C<br>C<br>C<br>C<br>C<br>C<br>C<br>C<br>C<br>C<br>C | R R 730 TTGTGAP V N 790 AGGACTC D S 850 CCTGCCP C Q 910 TGGACCCT  | S<br>CCT<br>CCT<br>L<br>GAT<br>M   | P<br>TGAG<br>E<br>GGGG<br>G<br>GTAG<br>Y                   | R<br>GGA<br>E<br>AAAA<br>N<br>CAA                                    | N GAT I ACAC T L ACCT L  | L CCC   | TCATO   | Y GCT L ACA H GTC S  | F<br>GCT<br>L<br>.CAI<br>I<br>.CCTF<br>Y   | G CAN I  | E TGA E CAT I ACAG R   | H<br>GCA<br>H<br>CCT<br>L<br>SACA<br>H    | P 77 ATGG G 8 CCCP Q 8 ATGG  | L<br>FACO<br>A<br>A<br>AGO<br>AGO<br>P<br>B<br>B<br>B<br>B<br>B<br>B<br>B<br>B<br>B<br>B<br>B<br>B<br>B<br>B<br>B<br>B<br>B<br>B   | S<br>CTGA<br>D<br>CCAA<br>N<br>ACCA<br>H             | F<br>I<br>I<br>CAA<br>K<br>CCI<br>L  | A<br>CCCG<br>R<br>LAAC<br>T  |
| A TGC A GGC A CTT F                                   | F<br>C<br>C<br>C<br>C<br>C<br>C<br>C<br>C<br>C<br>C<br>C | R R 730 TTGTGAP V N 790 AGGACTC D S 850 CCTGCCP C Q 910 TGGACCCT  | S<br>CCT<br>CCT<br>L<br>GAT<br>M   | P<br>TGAG<br>E<br>GGGG<br>G<br>GTAG<br>Y                   | R<br>GGA<br>E<br>AAAA<br>N<br>CAA                                    | N GAT I ACAC T L ACCT L  | L CCC   | TCATO   | Y GCT L ACA H GTC S  | F<br>GCT<br>L<br>.CAI<br>I<br>.CCTF<br>Y   | G CAN I  | E TGA E CAT I ACAG R   | H<br>GCA<br>H<br>CCT<br>L<br>SACA<br>H    | P 77 ATGG G 8 CCCP Q 8 ATGG  | L<br>FACO<br>A<br>A<br>AGO<br>AGO<br>P<br>B<br>B<br>B<br>B<br>B<br>B<br>B<br>B<br>B<br>B<br>B<br>B<br>B<br>B<br>B<br>B<br>B<br>B   | S<br>CTGA<br>D<br>CCAA<br>N<br>ACCA<br>H             | F<br>I<br>I<br>CAA<br>K<br>CCI<br>L  | A<br>CCCG<br>R<br>LAAC<br>T  |
| TGC A GGC A CTT F                                     | F CCTG C CCCP Q TTGG A                                   | R R 730 STGTGAP V N 790 AGGACTC D S 850 CCTGCCZ C Q 910 TGGACCT D L 970   | S<br>ACAGO<br>S<br>CCCTO<br>L<br>AGATO<br>M<br>CCGTO<br>V  | P TGAM E GGGG G GTAM Y GCCC                                | R<br>GGA<br>E<br>AAAA<br>N<br>CAA<br>N                               | GAT  I  CACCT  L  ATCF   | L CCC   | TCATO TGCG R 810 TGTT L 870 TGCT L 930 AGGG G 990   | Y GCT L ACA H GTC S TCI L  | F<br>GCI<br>L<br>CAI<br>I<br>CCI<br>Y  | G CAN I CCCI L CCCI D CCCC   | E TGA E CAT I ACAG R TTTT  | H AGCP H CCCT L FCAP H CCAP K             | ACCO P 77 ATGG G 8 8 CCCA ATGG G G S AGCT L 10   | ETTI<br>L<br>770<br>BASC<br>A<br>B30<br>P<br>B30<br>D<br>CGC<br>A<br>CGC<br>A  | S  CCAA  N  ACCA  H  CTGG                            | F<br>CAT<br>I<br>ACAA<br>K<br>ACCT<br>L<br>CAGGI<br>V  | A<br>CCCG<br>R<br>AAAC<br>T<br>CGCA<br>Q   |
| A TGC A   | F CCCTC C C CCCP Q TTGC A CCCCT                          | R R 730 TGTGAP V N 790 AGGACTC D S 850 CCTGCCZ C Q 910 TGGACCT D L 970 AGACTGT  | S ACAGO S CCCTO L AGATO M TCGTO V  | P TGAM E GGGG G GTAM Y GCCC P                              | R GGA E AAAA N CAA N CAA N TCA                                       | GAT I ACAC T ACCT L ATCP H AGCP  | L CCG V CAG V CAG V CAG ACCC  | ICATORIA TO   | Y GCT L ACA H GTC S TCT L  | F<br>GCI<br>L<br>CAI<br>I<br>CCT/<br>Y<br>CAC<br>T   | G CAN I L L L L L L L L L L L L L L L L L L                          | E TGA E CAT I ACAG R TTTT F  | H AGCA H CCCT L BACA H CCAA               | CCC<br>P 77<br>TGG<br>G 8<br>CCCP<br>Q 8<br>SATGG<br>G 9<br>SAGCT<br>L 10  | FTTT<br>L 770<br>FASC<br>A 330<br>AGCC<br>P 1990<br>FGGG<br>A 2010   | S CCAA D CCAA N ACCA R CTGG                          | F CAT I CAA K CCCI L GAGGI V   | A CCCG R AAAC T CCCA Q CCGA E CCTA   |
| A TGC A   | F CCCTC C C CCCP Q TTGC A CCCCT                          | R R 730 TGTGAP V N 790 AGGACTC D S 850 CCTGCCZ C Q 910 TGGACCT D L 970 AGACTGT  | S ACAGO S CCCTO L AGATO M TCGTO V  | P TGAM E GGGG G GTAM Y GCCC P                              | R GGA E AAAA N CAA N CAA N TCA                                       | GAT I ACAC T ACCT L ATCP H AGCP  | L CCG V CAG V CAG V CAG ACCC  | ICATORIA TO   | Y GCT L ACA H GTC S TCT L  | F<br>GCI<br>L<br>CAI<br>I<br>CCT/<br>Y<br>CAC<br>T   | G CAN I L L L L L L L L L L L L L L L L L L                          | E TGA E CAT I ACAG R TTTT F  | H AGCA H CCCT L BACA H CCAA               | CCC<br>P 77<br>TGG<br>G 8<br>CCCP<br>Q 8<br>SATGG<br>G 9<br>SAGCT<br>L 10  | FTTT<br>L 770<br>FASC<br>A 330<br>AGCC<br>P 1990<br>FGGG<br>A 2010   | S CCAA D CCAA N ACCA R CTGG                          | F CAT I CAA K CCCI L GAGGI V   | A CCCG R AAAC T CCCA Q CCGA E CCTA   |
| A TGC A   | F CCTG C C CCCP Q TTGC A CCCT L GTAI                     | R R 730 STGTGAP V N 790 AGGACTC D S 850 CCTGCCC C Q 910 TGGACCT D L 970 ACACTGT T V 1030  | S ACAGO S CCCTM AGATM V TCGTM V  | P TGAM E GGGG G GTAM Y GCC P GTTT F                        | R GGA E AAAA N CAA N CAA N TCA                                       | N GAT I CAC T ACCT L ATCA H AGCA H   | L CGC V CACC L CACC L   | TCAT  T 750 TGCG R 810 TGTT L 870 TGCT L 930 AGGG G 990 TGAT M 1050   | Y GCT L ACA H GTC S TCT L  | F<br>GCT<br>L<br>CAI<br>I<br>CCT<br>Y<br>CAC<br>T  | G CCCI   | E CAT I ACAG R CTTT F .  | H AGCA H CCCI L GACA H CAAGCA H           | ACCO P 77 ATGG G G CCCA ATGG G G ATGG G T AGCA T 10  | ETTT L 270 EAGO A 330 AGCO P 390 GGGGZ D 100 CCC2 Q  | S CTGA D CCCAA N ACCCA H CTGGG G AGTG                | F CAT I CAA K CCT L GAGT V GGGAC T   | A CCCG R AAAC T CGCA Q CGGA B CGTA Y   |
| A TGC A   | F CCTC C CCCP Q TTGC A CCCCT L GTAM                      | R R 730 TGTGAP V N 790 AGGACTC D S 850 CCTGCCF C Q 910 TGGACCT D L 970 ACACTGT T V 1030 CACTGAC                                     | S ACAGO S CCCTO L AGATO V TGATO M CCCTC  | P TGAM E GGGG G GTAM Y GCCO P GTT F                        | R GGA E AAA N CAA N CAA O TCA TCA                                    | N GAT I CAC T L CCT H AGC! H   | L CGC V CAGC V CAGC V CACC L ACCC L   | TCAT  T 750 TGCG R 810 TGCT L 870 TGCT L 930 AGGG G 990 TGAT M 1050 ACCT                                      | Y GCT L ACA H GTC S TCT L GCF  | F<br>GCI<br>L<br>.CAI<br>I<br>.CCI<br>Y<br>.CAC<br>T<br>.AGAM<br>K   | G CCAT I CCCT L ACGA P AGCC R  | E TGA E CAT I ACAG R CTTT F . GGAA K   | H AGCA H CCCI H CCCI H AGCA H ACCI O      | ACCO P 77 ATGG G G G CCCA AGCA T 10 CCCTC  | ETTT L 270 EAGO A 330 A 330 F 390 F 390 F 390 F 300 C C C C C C C C C C C C C C C C C C  | S TTGA D CCAA N ACCA H CTGG G G AGTG                 | F  CAT  I  CAP  K  CCT  L  GGGAC  T  | A CCCG R AAAC T CGCA Q CGGA B CCGTA Y  |
| A TGC A   | F CCTC C CCCP Q TTGC A CCCCT L GTAM                      | R R 730 TGTGAP V N 790 AGGACTC D S 850 CCTGCCF C Q 910 TGGACCT D L 970 ACACTGT T V 1030 CACTGAC                                     | S ACAGO S CCCTO L AGATO V TGATO M CCCTC  | P TGAM E GGGG G GTAM Y GCCO P GTT F                        | R GGA E AAA N CAA N CAA O TCA TCA                                    | N GAT I CAC T L CCT H AGC! H   | L CGC V CAGC V CAGC V CACC L ACCC L   | TCAT  T 750 TGCG R 810 TGCT L 870 TGCT L 930 AGGG G 990 TGAT M 1050 ACCT                                      | Y GCT L ACA H GTC S TCT L GCF  | F<br>GCI<br>L<br>.CAI<br>I<br>.CCI<br>Y<br>.CAC<br>T<br>.AGAM<br>K   | G CCAT I CCCT L ACGA P AGCC R  | E TGA E CAT I ACAG R CTTT F . GGAA K   | H AGCA H CCCI H CCCI H AGCA H ACCI O      | ACCO P 77 ATGG G G G CCCA AGCA T 10 CCCTC  | ETTT L 270 EAGO A 330 A 330 F 390 F 390 F 390 F 300 C C C C C C C C C C C C C C C C C C  | S TTGA D CCAA N ACCA H CTGG G G AGTG                 | F  CAT  I  CAP  K  CCT  L  GGGAC  T  | A CCCG R AAAC T CGCA Q CGGA B CCGTA Y  |
| A TGGGA A CTTT F GCCC G G G G G G                     | F CCTC C C CCCP Q TTGC A L CCCT I D GTAM                 | R R 730 TGTGAP V N 790 AGGACTC D S 850 CCTGCCA C Q 910 TGGACCT D L 970 ACACTGA T V 1030 CACTGAC L T 1090                            | S CCCTC L AGATT V TCGTT M CCCTC S  | P TGAM E GGGG G GTAM Y GCCC P GTT F                        | R GGA E AAAA N CAA N TCA TCA TCA TCA TCA                             | N GAT I CACCT L L ATCF H AGCF H  | L CCG V CAG V CAG L ACC L ACC L   | 750 TGAT  1 750 TGCG  R 810 TGTT  L 870 TGCT  L 930 AGGG  G 990 TGAT  M 1050 ACCT  L 1110                     | Y GCT L ACA H GTC S TCT L GCF Q CAC                                  | F<br>GCI<br>L<br>CAI<br>I<br>CCT!<br>Y<br>CAC<br>T<br>K<br>K<br>CAG!   | G CAT I CCCI L L CCCC P AGCC R                                       | E TGA E CAT I ACAG R   | H AGCA H CCCT L H FCAA K AGCA H ACT S     | P 77 ATGG 6 8 CCCP Q 6 8 CCCP 10 ACACAC T 10 CCCTC S 11  | L 770 FASC P 1390  | S TTGA D CCAA N ACCA H CTTGG G W GGGA D              | F  CAT  I  CCAP  K  CCCI  L  CGGAC  T  ATGE  | A CCCG R AAAC T CGCA Q CGGA B CGTA Y AGCA Q  |
| A TGGGA A CTTT F GCCC G G G G G G                     | F CCTC C CCCP Q TTGC A L GGACC P                         | R R 730 TGTGAP V N 790 AGGACTC D S 850 CCTGCCA C Q 910 TGGACCT D L 970 ACACTGT T V 1030 CACTGAC L T 1090 TGCTGCG                    | S CCCTC L AGATT N TCGTT V TGAT M AGATT AACTT AAC | P TGAM E GGGG G GTAM Y GCCC P GTT F GAC T TAT              | R GGA E AAA N CAA N CAA TCA TCA TCA L CAA                            | N GAT I L CAC T L L CTC H H L CCT H H CCT H Y CCT Y  | L CCG V CAG V CGT L ACC L ACC L ACC CCA   | 750 TGAT  750 TGCG  R  810 TGTT  L  870 TGCT  L  930 AGGG  G  990 TGAT  M  1050 ACCT  L  1110 CCAA            | Y GCT L ACA H GTC S TCI CAC  | F<br>GCI<br>L<br>CAI<br>I<br>CCTI<br>Y<br>CAC<br>T<br>K<br>CAGI<br>E   | G CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC                               | E TGA E CAT I ACAG R CTTTI GGAA K TCGA   | H AGCA H CCT L CAACCA H ACT S             | P 77 ATGG 8 8 8 CCCP 2 8 AGCT 10 ACCCT S 11  | L 170 A 130  | S TTGA D CCAA N ACCA H CTGG G G GGGA TCCT TCCT TCCT  | F CAT I CAA K CCCI L CAG T ATGA E  | A CCCG R AAAC T CGCA Q CGGA B CGTA Y AGCA Q ACCA   |
| A TGGGA A CTTT F GCCC G G G G G G                     | F CCTC C CCCP Q TTGC A L GGACC P                         | R R 730 TGTGAP V N 790 AGGACTC D S 850 CCTGCCA C Q 910 TGGACCT D L 970 ACACTGA T V 1030 CACTGAC L T 1090                            | S CCCTC L AGATT N TCGTT V TGAT M AGATT AACTT AAC | P TGAM E GGGG G GTAM Y GCCC P GTT F GAC T TAT              | R GGA E AAA N CAA N CAA TCA TCA TCA L CAA                            | N GAT I L CAC T L L CTC H H L CCT H H CCT H Y CCT Y  | L CCG V CAG V CGT L ACC L ACC L ACC CCA   | 750 TGAT  750 TGCG  R  810 TGTT  L  870 TGCT  L  930 AGGG  G  990 TGAT  M  1050 ACCT  L  1110 CCAA            | Y GCT L ACA H GTC S TCI CAC  | F<br>GCI<br>L<br>CAI<br>I<br>CCTI<br>Y<br>CAC<br>T<br>K<br>CAGI<br>E   | G CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC                               | E TGA E CAT I ACAG R CTTTI GGAA K TCGA   | H AGCA H CCT L CAACCA H ACT S             | CCCTO  | L 170 AGC P 1890 PGGG A A CCC22 C A G C C A G C C A G C C A G C C A G C C A G C C A G C C A G C C A G C C A G C C A G C C C A G C C C A G C C C A G C C C A G C C C A G C C C A G C C C C  | S TTGA D CCAA N ACCA H CTGG G G GGGA TCCT TCCT TCCT  | F CAT I CAA K CCCI L CAG T ATGA E  | A CCCG R AAAC T CGCA Q CGGA B CGTA Y AGCA Q ACCA   |
| A TGC A CTT F GCCC G G G GTGC G S                     | F  CCCCP  Q  TTGC  A  CCCCT  L  GTAN  N  GACC  P         | R R 730 TGTGAP V N 790 AGGACTC D S 850 CCTGCCF C Q 910 TGGACCT D L 970 ACACTGT T V 1030 CACTGAC L T 1090 TGCTGCF L E 1150           | S ACAGE S CCCTC L AGATT V TGAT M CCTC S AAACT L  | P TGAM E GGGG G GTAM Y GCCC P GTT F TAT I                  | R GGA AAA N CAA N CAA N TCA      | N GAT I CACCI L ATCE H ACCI H Y CTI Y  | L CGC V V CGT L ACCC L ACCC T   | TCAT'  1 750 PGCG R 810 PGTT L 870 PGCT L 930 AGGG G 990 AGCGT M 1050 ACCT L 1110 CCAA K 1170                 | GCT L GCF CAC  | F GCT L CAN I CAG K K CAG K R  | G CCAT I   | E  CAT  CAT  R  CTTT  F  GGAX  K  CCGX  AGGGA  AGGGA  A  | H AGCA  AGCA  H ACTY  R                   | CCCP P 77 ATGG G 88 CCCP Q 88 ATGGG G G 10 ACAC T 10 CCCTC S 11 ACCCC S 11 ACCCC CCCTC CCTC C | L 170 A 130  | S TTGA D CCAA N ACCA H CTGG G G GGGI D TCCT L        | F ACAT I ACAA K ACCT L GAGT V ATGA E D   | A CCG R AAAC T CGCA Q CGGA E CGTA Y AGCA Q ACCA Q  |
| A TGC A CTT F GCCC G G GTGC G GTGC G G G G G G G G G  | F  CCCTC C  C  C  C  C  C  C  C  C  C  C                 | R R 730 TGTGAP V N 790 AGGACTC D S 850 CCTGCCF C Q 910 TGGACCT T V 1030 CACACTGAC L T 1090 TGCTGGG L E 1150 CGGTGA                  | S ACAGE S CCCTG L AGAT V TCGT V TGAT M CCCTC S AAACT L AGGGA   | P TGAM E GGGG G GTAM Y GCCC P GTTT F TATT I                | R GGA E AAAA N CAA N CAA TCA TCA TCA TCA TCA TCA TCA TCA TCA         | N GAT I CACCO T L CACCO H H CCT! Y TCACCO T T T TCACCO T T T TCACCO T T T T T T T T T T T T T T T T T T   | L CGC V CAGC L CCA T CCA T CCA T CCA T  | TCATT  1 750 PGCG R 810 PGTT L 870 PGGT L 9300 AGGG G 90 PGAT M 1050 ACCT L 1110 CCAA K 1170 TCAA             | Y GCT L ACA H GTC S TCT CAC T GGA GGA GGA GGA GGA GGA GGA GGA GGA G  | F GCT L CAT I CAT Y CAC K K CAGA R CA | G CCAT I CCCT L ACGA P AGCC R I AGCC AGGA I AGGC AGGC AGGC AGGC AGGC | E TGAT  I ACAG  R TTTT  F GGAN  K TCGN  AGGGG  | H AGCA  AGCA  H ACTO  R ACGG              | CCCCP P 77 ATGG G 8 8 CCCP Q E ATGG G G G G G G G G G G G G G G G G G  | L TTI  | S TTGA D D TTGA N ACCA H CTTGG G AGTG D L CGTA       | F CATE  ACCT  L CARGO  V ACCT  T ATGM  ACCT  ATGM  ACCT  ACC | A CCCA T CCCA Q CCCA Y ACCA Q CCCA Q CCCTGCA Q CCCTGCA Q CCCTGCA CCCA CCCTGCA CCCA CCCTGCA CCCA CC |
| A TGC A CTT F GCCC G G GTGC G GTGC G G G G G G G G G  | F  CCCTC C  C  C  C  C  C  C  C  C  C  C                 | R R 730 TGTGAP V N 790 AGGACTC D S 850 CCTGCCF C Q 910 TGGACCT D L 970 ACACTGT T V 1030 CACTGAC L T 1090 TGCTGCF L E 1150           | S ACAGE S CCCTG L AGAT V TCGT V TGAT M CCCTC S AAACT L AGGGA   | P TGAM E GGGG G GTAM Y GCCC P GTTT F TATT I                | R GGA E AAAA N CAA N CAA TCA TCA TCA TCA TCA TCA TCA TCA TCA         | N GAT I CACCO T L CACCO H H CCT! Y TCACCO T T T TCACCO T T T TCACCO T T T T T T T T T T T T T T T T T T   | L CGC V CAGC L CCA T CCA T CCA T CCA T  | TCATT  1 750 PGCG R 810 PGTT L 870 PGGT L 9300 AGGG G 90 PGAT M 1050 ACCT L 1110 CCAA K 1170 TCAA             | Y GCT L ACA H GTC S TCT CAC T GGA GGA GGA GGA GGA GGA GGA GGA GGA G  | F GCT L CAT I CAT Y CAC K K CAGA R CA | G CCAT I CCCT L ACGA P AGCC R I AGCC AGGA I AGGC AGGC AGGC AGGC AGGC | E TGAT  I ACAG  R TTTT  F GGAN  K TCGN  AGGGG  | H AGCA  AGCA  H ACTO  R ACGG              | CCCCP P 77 ATGG G 8 8 CCCP Q E ATGG G G G G G G G G G G G G G G G G G  | L TTI  | S TTGA D D TTGA N ACCA H CTTGG G AGTG D L CGTA       | F CATE  ACCT  L CARGO  V ACCT  T ATGM  ACCT  ATGM  ACCT  ACC | A CCCA T CCCA Q CCCA Y ACCA Q CCCA Q CCCTGCA Q CCCTGCA Q CCCTGCA CCCA CCCTGCA CCCA CCCTGCA CCCA CC |
| A TGC A CTT F GCCC G G GTGC G TGC G TGC TGC TGC TGC T | F CCCCP CCCP Q TTGC A CCCCP A CCCCP L CCCCCP P           | R R 730 TGTGAP V N 790 AGGACTC D S 850 CCTGCCF C Q 910 TGGACCT T V 1030 CACACTGA L T 1090 TGCTGGE L E 1150 CGGTGAF V K 1210         | S ACAGE S CCCTG L AGAT V TCGT V TGAT M CCCTC S AACT L AGGGA  | P TGAM E GGGi G GTAM Y GCCC T TAT I AGCT L                 | R GGA E AAAA N CAAA N CAAA N TCA | N GAT I CACCT L L ATCA H Y CTAC T T CACCT T T  | L CCC   | TCAT'  1 750 PGCG R 810 PGTT L 870 PGCT L 930 AGGG G 990 TGAT M 1050 ACCT L 1110 CCAA K 1170 TCAA             | GCT<br>L<br>ACA<br>H<br>GTC<br>S<br>TCI<br>CGA<br>TCA<br>CGA<br>K    | F GCT L CAT I CCTF Y CAC T AGAM K CAGM K CAGM K  | G CCAT I CCCI L ACGI P AGCCI R AGACI E AGCCI R                       | E TGAT E CAT I ACAG R TTTT F CGGA K CCGA AGGGG A GGGT Y  | H AGCA H CCT K AGCA H ACT CCT CC R ACGG G | CCCCP  77 ATGG  6 8 8 CCCAP Q 8 8 CCCAP Q 8 8 CCCAP Q 10 10 10 10 10 10 10 10 10 10 10 10 10   | 770 FASC P 1990 CASC P 1990 CA | S TTGA D CCAA N ACCA H CTGG G AGTG U CGGI            | F ACAT I I ACAM K ACCI L O GGAC T ATGM E F   | A CCCA T CCCA B CCCTA Y ACCCA Q CCCTA C C C C C C C C C C C C C C C C C                            |
| A TGC A CTT F GCCC G G GTGC G TGC G TGC TGC TGC TGC T | F F CCCTC C C C C C C C C C C C C C C C                  | R R 730 TGTGAP V N 790 AGGACTC D S 850 CCTGCCF C Q 910 TGGACCT T V 1030 CACACTGA L T 1090 TGCTGGC L E 1150 CGGTGAC V K 1210 TGGGTGA | S CCCTC L AGAT V TGAT M CCTC S AGGA ACT L AGGAA E CCAT   | P TGAM E GGGG G GTAM Y GCCC P GTT F GAC T TAT I GCT L CATA | R GGA E AAAA N CAAA N CAAA N TCA | N GAT I CACCT L L CCT H H AGCF T T CACCT T T C | L CCG V AGG V AGG L ACC L ACC L TGT TGT TGT TGT   | TCATT  1 750 PGCG R 810 PGTT L 870 PGCT L 930 AGGG G 990 TGAT M 1050 ACCT L 1110 CCAA K 1170 TCAA K 1230 ACAT | Y GCT L GTC S TCT L GCF CAC T GAM CGAM CCAC W                        | F GCT L CAT I CCT Y CAC T AGA K CAGA K CAGA K CCT CCT CCT CCT CCT CCT CCT CCT CCT C  | G CAT  | E TGAT E CATTI ACAG R CTTTI F GGAA K CTTTI Y CGAT T T CGA | H AGCA H CCTC R ACCC R ACCC CCA           | P 77 ATGG G 88 CCCA G G S 10 ACCA G G S 11 ACCA G G G G G G G G G G G G G G G G G G  | 770 FASC P 1990 PGGG P 250 PGG P 250 PG | S CCAA N ACCA N H CTGG G AGTG V CGGT L CGGT Y        | F CAT I CAA K K CCT L GGGAC T T CGGAC F F CTZ  | A CCG R AAAC T CGCA Q CGGA B CGTA Y AGCA Q ACCA C  |
| A TGC A CTT F GCCC G G GTGC G TGC G TGC TGC TGC TGC T | F F CCCTC C C C C C C C C C C C C C C C                  | R R 730 TGTGAP V N 790 AGGACTC D S 850 CCTGCCF C Q 910 TGGACCT T V 1030 CACACTGA L T 1090 TGCTGGE L E 1150 CGGTGAF V K 1210         | S CCCTC L AGAT V TGAT M CCTC S AGGA ACT L AGGAA E CCAT   | P TGAM E GGGG G GTAM Y GCCC P GTT F GAC T TAT I GCT L CATA | R GGA E AAAA N CAAA N CAAA N TCA | N GAT I CACCT L L CCT H H AGCF T T CACCT T T C | L CCG V AGG V AGG L ACC L ACC L TGT TGT TGT TGT   | TCATT  1 750 PGCG R 810 PGTT L 870 PGCT L 930 AGGG G 990 TGAT M 1050 ACCT L 1110 CCAA K 1170 TCAA K 1230 ACAT | Y GCT L GTC S TCT L GCF CAC T GAM CGAM CCAC W                        | F GCT L CAT I CCT Y CAC T AGA K CAGA K CAGA K CCT CCT CCT CCT CCT CCT CCT CCT CCT C  | G CAT  | E TGAT E CATTI ACAG R CTTTI F GGAA K CTTTI Y CGAT T T CGA | H AGCA H CCTC R ACCC R ACCC CCA           | P 77 ATGG G 88 CCCA G G S 10 ACCA G G S 11 ACCA G G G G G G G G G G G G G G G G G G  | 770 FASC P 1990 PGGG P 250 PGG P 250 PG | S CCAA N ACCA N H CTGG G AGTG V CGGT L CGGT Y        | F CAT I CAA K K CCT L GGGAC T T CGGAC F F CTZ  | A CCG R AAAC T CGCA Q CGGA B CGTA Y AGCA Q ACCA C  |
| A TGGGGA A CTTT F GGGG G G GGGG G G GTGGG G G G G G   | F F CCCTC C C C C C C C C C C C C C C C                  | R R 730 TGTGAP V N 790 AGGACTC D S 850 CCTGCCF C Q 910 TGGACCT T V 1030 CACACTGA L T 1090 TGCTGGC L E 1150 CGGTGAC V K 1210 TGGGTGA | S CCCTC L AGAT V TGAT M CCTC S AGGA ACT L AGGAA E CCAT   | P TGAM E GGGG G GTAM Y GCCC P GTT F GAC T TAT I GCT L CATA | R GGA E AAAA N CAAA N CAAA N TCA | N GAT I CACCT L L CCT H H AGCF T T CACCT T T C | L CCC. V V CCC. L CCC. T CCC. | TCATT  1 750 PGCG R 810 PGTT L 870 PGCT L 930 AGGG G 990 TGAT M 1050 ACCT L 1110 CCAA K 1170 TCAA K 1230 ACAT | GCT<br>L<br>ACA<br>H<br>GTC<br>S<br>TCI<br>L<br>GCF<br>Q<br>CAC<br>K | F GCT L CAT I CCT Y CAC T AGA K CAGA K CAGA K CCT CCT CCT CCT CCT CCT CCT CCT CCT C  | G CAT  | E TGAT E CATTI ACAG R CTTTI F GGAA K CTTTI Y CGAT T T CGA | H AGCA H CCTC R ACCC R ACCC CCA           | P 77 ATGG G 8 E E E E E E E E E E E E E E E E E E E  | 770 FASC P 1990 PGGG P 250 PGG P 250 PG | S TTGA D CCCAM H ACCAM H CTGG G AGTG D CGTM Y GCAM I | F CAT I CAA K K CCT L GGGAC T CGGAC T F CTZ  | A CCG R AAAC T CGCA Q CGGA B CGTA Y AGCA Q ACCA C  |

CCCCTCAAGCCCAGGACCAATAACCGCACAAGCCCCCGGGACAACACCCTCTTACAGCA PLKPRTNNRTSPRDNTLLQQ 1370 1330 1350 GAAGCTACTTCAGGAAGCCTACGTGACCCCTAAGGACGATATCCGGCTGGTCGGGGAGCT K L L Q E A Y V T P K D D I R L V G E L 1410 1430 1390 GGTGACTGTCATTGGGGCTATCATCATCCTGCTGGTAGAGGTTCCAGACATCTTCAGAAT V T V I G A I I I L L V E V P D I F R M 1470 1490 GGGGGTCACTCGCTTCTTTGGACAGACCATCCTTGGGGGCCCATTCCATGTCCTCATCAT G V T R F F G Q T I L G G P F H V L I I 1550 1530 1510 CACCTATGCCTTCATGGTGCTGGTGACCATGGTGATGCGGCTCATCAGTGCCAGCGGGGA T Y A F M V L V T M V N R L I S A S G E 1570 1590 1610 GGTGGTACCCATGTCCTTTGCACTCGTGCTGGGGTGCAACGTCATGTACTTCGCCCG V V P M S F A L V L G W C N V M Y F A R 1670 1650 AGGATTCCAGATGCTAGGCCCCTTCACCATCATGATTCAGAAGATGATTTTTGGCGÄCCT G F Q M L G P F T I M I Q K M I F G D L 1710 1690 GATGCGATTCTGCTGGCTGATGGCTGTGGTCATCCTGGGCTTTGCTTCAGCCTTCTATAT MRFCWLMAVVILGFASAFYI 1790 1770 CATCTTCCAGACAGAGGACCCCGAGGAGCTAGGCCACTTCTACGACTACCCCATGGCCCT I F Q T E D P E E L G H F Y D Y P M A L 1830 1810 1850 GTTCAGCACCTTCGAGCTGTTCCTTACCATCATCGATGGCCCAGCCAACTACAACGTGGA FSTFELFLTIIDGPANYNVD 1890 1910 1870 CCTGCCCTTCATGTACAGCATCACCTATGCTGCCTTTGCCATCATCGCCACACTGCTCAT L P F M Y S I T Y A A F A I I A T L L M 1950 1970 GCTCAACCTCCTCATTGCCATGATGGGCGACACTCACTGGCGAGTGGCCCCATGAGCGGGA L N L L I A M M G D T H W R V A H E R D 2010 2030 TGAGCTGTGGAGGGCCCAGATTGTGGCCACCACGGTGATGCTGGAGCGGAAGCTGCCTCG E L W R A Q I V A T T V M L E R K L P R 2070 2090 CTGCCTGTGGCCTCGGGGATCTGCGGACGGGAGTATGGCCTGGGGGACCGCTGGTT C L W P R S G I C G R E Y G L G D R W F 2150 2110 2130 CCTGCGGGTGGAAGACAGGCAAGATCTCAACCGGCAGCGGATCCAACGCTACGCACAGGC 2190 2170 CTTCCACACCCGGGGCTCTGAGGATTTGGACAAAGACTCAGTGGAAAAACTAGAGCTGGG F H T R G S E D L D K D S . V E K L E L G 2250 2270 CTGTCCCTTCAGCCCCCACCTGTCCCTTCCTACGCCCTCAGTGTCTCGAAGTACCTCCCG C P F S P H L S L P T P S V S R S T S R 2310 2290 CAGCAGTGCCAATTGGGAAAGGCTTCGGCAAGGGACCCTGAGGAGACCCTGCGTGGGAT S S A N W E R L R Q G T L R R D L R G I 2370 2390 AATCAACAGGGGTCTGGAGGACGGGGAGAGCTGGGAATATCAGATCTGACTGCGTGTTCT I N R G L E D G E S W E Y Q I 2430 2510 2490 AACACCCAGAGGTCTCATCTCCCAGGCCCCAGGGAGAAAGAGGAGTAGCATGAACGCCAA 2550 2570 GGAATGTACGTTGAGAATCACTGCTCCAGGCCTGCATTACTCCTTCAGCTCTGGGGCAGA

Fig. 7 / continuation 2

| 2590                | 2610                 | 2630                   |
|---------------------|----------------------|------------------------|
| GGAAGCCCAGCCCAAGCAC | XGGGCTGGCAGGGCGTGAG  | GAACTCTCCTGTGGCCTGCTCA |
| 2650                | 2670                 | 2690                   |
| TCACCCTTCCGACAGGAGG | CACTGCATGTCAGAGCACTT | TAAAAACAGGCCAGCCTGCTTG |
| 2710                | 2730                 | 2750                   |
| GGCCCTCGGTCTCCACCC  | CAGGGTCATAAGTGGGGAGA | GAGCCCTTCCCAGGGCACCCAG |
| 2770 .              | 2790                 | 2810                   |
| GCAGGTGCAGGGAAGTGC  | AGAGCTTGTGGAAAGCGTGT | GAGTGAGGGAGAACGGC      |
| 2830                | 2850                 | 2870                   |
| TCTGGGGGTGGGAAGTGG  | GCTAGGTCTTGCCAACTCC  | ATCTTCAATAAAGTCGTTTTCG |
| 2890                | 2910                 |                        |
| GATCCCTAAAAAAAAAAA  | AAAAAAAAAAAA         |                        |

MGLSLPKEKGLILCLWSKFCRWFQRRESWAQSRDEQNLLQOKRIWESPLLLAAKDNDVQALNKLLKYEDCKVHQRGAMGETALHIA ALYDNLBAAMVIMEAAPELVFEPMTSELYEGQTALHIAVVNQNMNLVRALLARRASVSARATGTAFRRSPRNLIYFGEHPLSFAAC VNSEEIVRLLIEHGADIRAQDSLGNTVLHILILQPNKTFACQMYNLLLSYDRHGDHLQPLDLVPNHQGLTPFKLAGVEGNTVMFQH LMQKRKHTQWTYGPLTSTLYDLTEIDSSGDEQSLLELIITTKKREARQILDQTPVKELVSLKWKRYGRPYFCMLGAIYLLYIICFT MCCIYRPLKPRTNNRTSPRDNTLLQQKLLQEAYVTPKDDIRLVGELVTVIGAIIILLVEVPDIFRMGVTRFFGQTILGGPFHVLII TYAFMVLVTMVMRLISASGEVVPMSFALVLGWCNVMYFARGFQMLGPFTIMIQKMIFGDLMRFCWLMAVVILGFASAFYIIFQTED PEELGHFYDYPMALFSTFELFLTIIDGPANYNVDLPFMYSITYAAFAIIATLLMLNLLIANMGDTHWRVAHERDELWRAQIVATTV MLERKLPRCLWPRSGICGREYGLGDRWFLRVEDRQDLNRQRIQRYAQAFHTRGSEDLDKDSVEKLELGCPFSPHLSLPTPSVSRST SRSSANWERLRQGTLRRDLRGIINRGLEDGESWEYQI

Figure 8:

A) . ATGGGTTTGTCACTGCCCAAGGAGAAAGGGCTAATTCTCT MGLSLPKEKGLILC 270 290 GCCTATGGAGCAAGTTCTGCAGATGGTTCCAGAGACGGGAGTCCTGGGCCCAGAGCCGAG LWSKFCRWFQRRESWAQSRD 310 330 350 ATGAGCAGAACCTGCTGCAGCAGAAGAGGGTCTCGGGAGTCTCCTCTCTAGCTGCCA E Q N L L Q Q K R I W E S P L L L A A K 370 390 410  ${\tt AAGATAATGATGTCCAGGCCCTGAACAAGTTGCTCAAGTATGAGGATTGCAAGGTGCACC}$ D N D V Q A L N K L L K Y E D C K V H Q 450 470 AGAGAGGAGCCATGGGGGAAACAGCGCTACACATAGCAGCCCTCTATGACAACCTGGAGG RGAMGETALHIAALYDNLEA 510 530  ${\tt CCGCCATGGTGCTGATGGAGGCTGCCCCGGAGCTGGTCTTTGAGCCCATGACATCTGAGC}$ AMVLMEAAPELVFEPMTSEL 570 590 TCTATGAGGGTCAGACTGCACTGCACATCGCTGTTGTGAACCAGAACATGAACCTGGTGC Y E G Q T A L H I A V V N Q N M N L V R 610 630  ${\tt GAGCCCTGCTTGCCGGGGGCCAGGGCCAGGGCCACAGGCCACAGGCCACTGCCTTCCGCC}$ A L L A R R A S V S A R A T G T A F R R 670 690  ${\tt GTAGTCCCTGCAACCTCATCTACTTTGGGGAGCACCCTTTGTCCTTTGCTGCCTGTGTGA}$ SPCNLIYFGEHPLSFAACVN

Fig. 8 / contin 11

770 ACAGTGAGGAGATCGTGCGGCTGCTCATTGAGCATGGAGCTGACATCCGGGCCCAGGACT S E E I V R L L I E H G A D I R A Q D S 830 810 LGNTVLHILILQPNKTFACQ 870  ${\tt AGATGTACAACCTGTTGCTGTCCTACGACAGACATGGGGACCACCTGCAGCCCCTGGACC}$ MYNLLLSYDRHGDHLQPLDL 950 930 910 V P N H Q G L T P F K L A G V E G N T V 1010 990 TGATGTTTCAGCACCTGATGCAGAAGCGGAAGCACCCCAGTGGACGTATGGACCACTGA 1070 1030 1050 CCTCGACTCTCTATGACCTCACAGAGATCGACTCCTCAGGGGATGAGCAGTCCCTGCTGG STLYDLTEIDSSGDEQSLLE 1090 1110 1130 AACTTATCATCACCACCAAGAAGCGGGAGGCTCGCCAGATCCTGGACCAGACGCCGGTGA LIITTKKREARQILDQTPVK 1170 1190 AGGAGCTGGTGAGCCTCAAGTGGAAGCGGTACGGGCGGCGGTACTTCTGCATGCTGGGTG E L V S L K W K R Y G R P Y F C M L G A 1230 1250 CCATATATCTGCTGTACATCATCTGCTTCACCATGTGCTGCATCTACCGCCCCCTCAAGC IYLLYIICFTNCCIYRPLKP 1290 1310 1270 CCAGGACCAATAACCGCACGAGCCCCCGGGACAACACCCTCTTACAGCAGAAGCTACTTC R T N N R T S P R D N T L L Q Q K L L Q 1370 1330 1350 AGGAAGCCTACATGACCCCTAAGGACGATATCCGGCTGGTCGGGGAGCTGGTGACTGTCA EAYMTPKDDIRLVGELVI 1390 1410 1430  ${\tt TTGGGGCTATCATCCTGCTGGTAGAGGTTCCAGACATCTTCAGAATGGGGGTCACTC}$ GAIIILLVEVPDIFRMGVTR 1470 1490  ${\tt GCTTCTTTGGACAGACCATCCTTGGGGGCCCATTCCATGTCCTCATCATCACCTATGCCT}$ FFGQTILGGPFHVLIITYAF 1510 1530 TCATGGTGCTGGTGACCATGGTGATGCGGCTCATCAGTGCCAGCGGGGAGGTGGTACCCA  $\begin{smallmatrix} M & V & L & V & T & M & V & M & R & L & I & S & A & S & G & E & V & V & P & M \\ \end{smallmatrix}$ 1570 1590 -1610 TGTCCTTTGCACTCGTGCTGGGCTGGTGCAACGTCATGTACTTCGCCCGAGGATTCCAGA S F A L V L G W C N V M Y F A R G F Q M 1650 1670 TGCTAGGCCCCTTCACCATCATGATTCAGAAGATGATTTTTGGCGACCTGATGCGATTCT LGPFTIMIQKMIFGDLMRFC 1710 1690 1730  ${\tt GCTGGCTGATGGCTGTGCTCATCCTGGGCTTTGCTTCAGCCTTCTATATCATCTTCCAGA}$ WLMAVVILGFASAFYIIFQT 1770 1790 CAGAGGACCCCGAGGAGCTAGGCCACTTCTACGACTACCCCATGGCCCTGTTCAGCACCT EDPEELGHFYDYPMALFSTF 1810 1830 1850 TCGAGCTGTTCCTTACCATCATCGATGGCCCAGCCAACTACAACGTGGACCTGCCCTTCA ELFLTIIDGPANYNVDLPFM 1910 1870 1890 TGTACAGCATCACCTATGCTGCCTTTGCCATCATCGCCACACTGCTCATGCTCAACCTCC Y S I T Y A A F A I I A T L L M L N L L TCATTGCCATGATGGGCGACACTCACTGGCGGAGTGGCCCATGAGCGGGATGAGCTGTGGA

Fig. 8 / conti: I A M M G D T H W R V A H E R D B L W R 2030 1990 2010 GGGCCCAGATTGTGGCCACCACGGTGATGCTGGAGCGGAAGCTGCCTCGCTGCCTGTGGC AQIVATTVMLERKLPRCLWP 2070 2090 CTCGCTCCGGGATCTGCGGACGGGAGTATGGCCTGGGAGACCGCTGGTTCCTGCGGGTGG R S G I C G R E Y G L G D R W F L R V E 2150 2110 21.30 AAGACAGGCAAGATCTCAACCGGCAGCGGATCCAACGCTACGCACAGGCCTTCCACACCC D R Q D L N R Q R I Q R Y A Q A F H T R 2170 2190 2210 CGGCTCTGAGGATTTGGACAAAGACTCAGTGGAAAAACTAGAGCTGGGCTGTCCCTTCA G S E D L D K D S V B K L E L G C P F S 2250 2270 GCCCCACCTGTCCCTTCCTATGCCCTCAGTGTCTCGAAGTACCTCCCGCAGCAGTGCCA P H L S L P M P S V S R S T S R S S A N 2290 2310 2330 ATTGGGAAAGGCTTCGGCAAGGGACCCTGAGGAGACCTGCGTGGGATAATCAACAGGG WERLROGTLRRDLRGIINRG 2370 2390 GTCTGGAGGACGGGGAGAGCTGGGAATATCAGATCTGA LEDGESWEYQI\*

MGLSLPKEKGLILCLWSKFCRWFQRRESWAQSRDEQNLLQQKRIWESPLLLAAKDNDVQALNKLLKYEDCKVHQRGAMGETALHIA
ALYDNLEAAMVLMEAAPELVFEPMTSELYEGQTALHIAVVNQNMNLVRALLARRASVSARATGTAFRRSPCNLIYFGEHPLSFAAC
VNSEEIVRLLIEHGADIRAQDSLGNTVLHILILQPNKTFACQMYNLLLSYDRHGDHLQPLDLVPNHQGLTPFKLAGVEGNTVMFQH
LMQKRKHTQWTYGPLTSTLYDLTEIDSSGDEQSLLELIITTKKREARQILDQTPVKELVSLKWKRYGRPYFCMLGAIYLLYIICFT
MCCIYRPLKPRTNNRTSPRDNTLLQQKLLQEAYMTPKDDIRLVGELVTVIGAIIILLVEVPDIFRMGVTRFFGQTILGGPFHVLII
TYAFMVLVTNVMRLISASGEVVPMSFALVLGWCNVMYFARGFQMLGPFTIMIQKMIFGDLMRFCWLMAVVILGFASAFYIIFQTED
PEELGHFYDYPMALFSTFELFLTIIDGPANYNVDLPFMYSITYAAFAIIATLLMLNLLIAMMGDTHWRVAHERDELWRAQIVATTV
MLERKLFRCLWPRSGICGREYGLGDRWFLRVEDRQDLNRQRIQRYAQAFHTRGSEDLDKDSVEKLELGCPFSPHLSLPMPSVSRST
SRSSANWERLRQGTLRRDLRGIINRGLEDGESWEYQI

B)

 $\tt CCTCTACAGGGAGACGGTGGGCCGGCCCTTGGGGGGGGGTGATGTGGCCCCAAGGCTGAGTCCCGTCAGGGTCTGGCCTCAGGCCTCAGGGCTCAGGGTCTGGGCCTCAGGGCTCAGGGCTCAGGGTCTGGGCCTCAGGGCTCAGGGCTCAGGGTCTGGCCCTCAGGGCTCAGGCTCAGGGCTCAGGGCTCAGGGCTCAGGGCTCAGGCTCAGGCTCAGGGCCTCAGGGCTCAGGGCTCAGGCTCAGGGCTCAGGGC$ GGCCCCCAAGGAGCCGGCCCTACACCCCATGGGTTTGTCACTGCCCAAGGAGAAAGGGCTAATTCTCTGCCTATGGAGCAAGTTCT GCAGATGGTTCCAGAGACGGGAGTCCTGGGCCCAGAGCCGAGATGAGCAGAACCTGCTGCAGCAGAAGAGGATCTGGGAGTCTCCT CATGGGGGAAACAGCGCTACACATAGCAGCCCTCTATGACAACCTGGAGGCCGCCATGGTGCTGATGGAGGCTGCCCCGGAGCTGG TCTTTGAGCCCATGACATCTGAGCTCTATGAGGGTCAGACTGCACTGCACATCGCTGTTGTGAACCAGAACATGAACCTGGTGCGA GCCCTGCTTGCCCCCAGGGCCAGTGTCTCTGCCAGAGCCACAGGCACTGCCTTCCGCCGTAGTCCCCGCAACCTCATCTACTTTGG GGAGCACCCTTTGTCCTTTGCTGCCTGTGTGAACAGTGAGGAGATCGTGCGGCTGCTCATTGAGCATGGAGCTGACATCCGGGCCC TGCAGAAGCGGAAGCACACCCAGTGGACGTATGGACCACTGACCTCCGACTCTCTATGACCTCACAGAGATCGACTCCTCAGGGGAT GAGCAGTCCCTGCTGGAACTTATCATCACCACCAAGAAGCGGGGAGGCTCGCCAGATCCTGGACCAGACGCCGGTGAAGGAGCTGGT GAGCCTCAAGTGGAAGCGGTACGGGCGGCCGTACTTCTGCATGCTGGGTGCCATATATCTGCTGTACATCATCTGCTTCACCATGT GCTGCATCTACCGCCCCCTCAAGCCCAGGACCAATAACCGCACAAGCCCCCGGGACAACACCCTCTTACAGCAGAAGCTACTTCAG GAAGCCTACGTGACCCCTAAGGACGATATCCGGCTGGTCGGGGGGCGCTGTCATTGGGGGCTATCATCATCCTGCTGGTAGA GGTTCCAGACATCTTCAGAATGGGGGTCACTCGCTTCTTTGGACAGACCATCCTTGGGGGGCCCATTCCATGTCCTCATCACCCT GCCCTGTTCAGCACCTTCGAGCTGGTCCTTACCATCATCGATGGCCCAGCCAACTACAACGTGGACCTGCCCTTCATGTACAGCAT CGCTCCGGGATCTGCGGACGGGGTATGGCCTGGGGGACCGCTGCTTCCTGCGGGTGGAAGACAGGCAAGATCTCAACCGGCAGCG

Fig. 8 / continuation 3

c.)

CAAACTCACAGCCCTCTCCAAACTGGCTGGGGCTGCTGGGACACTCCCAAGGAACTCGTCAGGAAGGCAGGAGACACGGAACACGGA CCTCTACAGGGAGACGGTGGGCCGCCCCTTGGGGGGGCTGATGTGGCCCCAAGGCTGAGTCCCGTCAGGGTCTGGCCCTCA GCAGATGGTTCCAGAGACGGGAGTCCTGGGCCCAGAGCCGACATGAGCAGAACCTGCTGCAGCAGAAGAGGATCTGCGCAGTCTCCT CTCCTTCTAGCTGCCAAAGATAATGATGTCCAGGCCCTGAACAAGTTGCTCAAGTATGAGGATTGCAAGGTGCACCAGAGAGGAGG CATGGGGGAACAGCGCTACACATAGCAGCCCTCTATGACAACCTGGAGGCCGCCATGGTGCTGATGGAGGCTGCCCCGGAGCTGG TCTTTGAGCCCATGACATCTGAGCTCTATGAGGTCCTGACTGCCCATCACTTGAACGCCTGCCCCTGAAATGCCAGGGCCTAGAG AAGAGGAAGAGATGGGCAGCAGCTGGATCCCCTGGGAATCCTGAACACCCGAGAGCTCCCTGTTCTCCATCCCAGGCTACCCCTGA GGGAAAGAGACTGGGGTGCATATGGGAGGGACCCCCTGCAGGATCCTGGGGACAGACCCGTGACAGCTGTCTCTGGGCCAGG GAGCCACAGGCACTGCCTTCCGCCGTAGTCCCTGCAACCTCATCTACTTTGGGGAGCACCCTTTGTCCTTTGCTGCCTGTGTGAAC AGATGTACAACCTGTTGCTGTCCTACGACAGACATGGGGACCACCTGCAGCCCCTGGACCTCGTGCCCAATCACCAGGGTCTCACC CCTTTCAAGCTGGCTGGAGTGGAGGGTAACACTGTGATGTTTCAGCACCTGATGCAGAAGCGGAAGCACACCCCAGTGGACGTATGG ACCACTGACCTCGACTCTCTATGACCTCACAGAGATCGACTCCTCAGGGGGATGAGCAGTCCCTGCTGGAACTTATCATCACCACCA TTCTGCATGCTGGGTGCCATATATCTGCTGTACATCATCTGCTTCACCATGTGCTGCATCTACCGCCCCCTCAAGCCCAGGACCAA TAACCGCACGAGCCCCCGGGACAACACCCTCTTACAGCAGAAGCTACTTCAGGAAGCCTACATGACCCCTAAGGACGATATCCGGC TGGTCGGGGAGCTGGTGACTGTCATTGGGGCTATCATCATCCTGCTGGTAGAGGTTCCAGACATCTTCAGAATGGGGGTCACTCGC TTCTTTGGACAGACCATCCTTGGGGGCCCATTCCATGTCCTCATCATCACCTATGCCTTCATGGTGATGACCATGGTGATGCG GCTCATCAGTGCCAGCGGGAGGTGGTACCCATGTCCTTTGCACTCGTGCTGGGCTGCTAACGTCATGTACTTCGCCCGAGGAT ATCCTGGGCTTTGCTTAGACAGAGGAGCCCGAGGAGCTAGGCCACTTCTACGACTACCCCATGGCCCTGTTCAGCACCTTCGAGCT GGTCCTTACCATCATCGATGGCCCAGCCAACTACAACGTGGACCTGCCCTTCATGTACAGCATCACCTATGCCGTTTGCCATCA GTATGCCTGGEAGACCCCTGGTTCCTGCGCGTGGAAGACAGGCAAGATCTCAACCGGCAGCGGATCCAACGCTACGCACAGGCCT TCCACACCCGGGGCTCTGAGGATTTGGACAAAGACTCAGTGGAAAAACTAGAGCTGGGCTGTCCCTTCAGCCCCCACCTGTCCCTT CCTATGCCCTCAGTGTCTCGAAGTACCTCCCGCAGCAGTGCCAATTGGGAAAGGCTTCGGCAAGGGACCCTGAGGAAAGACCTGCG TGGGATAATCAACAGGGGTCTGGAGGACGGGGAGAGCTGGGAATATCAGATCTGACTGCGTGTTCTCACTTCGCTTCCTGGAACTT GCTCTCATTTTCCTGGGTGCATCAAACAAAACAAAAACCAAACACCCAGAGGTCTCATCTCCCAGGCCCCCAGGGAAAAGAGGGGGT AGCATGAACGCCAAGGAATGTACGTTGAGAATCACTGCTCCAGGCCTGCATTACTCCTTCAGCTCTGGGGCAGAGGCACAGCC CAAGCACGGGGCTGGCAGGACCTCTCCTGTGGCCTGCTCATCACCCTTCCGACAGGAGCACTGCATGTCAGAGCACTT TAAAAACAGGCCAGCCTGCTTGGGCCCTCGGTCTCCACCCCAGGGTCATAAGTGGGGAGAGACCCCTTCCCAGGGCACCCAGGCAG GTGCAGGGAGTGCAGAGCTTGTGGAAAGCGTTGTGAGTCAGGGAGACAGGAACGGCTCTGGGGGTGGGAAGTGGGGCTAGGTCTTG 

D.)

### Fig. 8 / continuation ·

GTGATGTTTCAGCACCTGATGCAGAAGCGGAAGCACCCCAGTGGACGTATGGACCTCACTGACTCTCTATGACCTCACAGA GATCGACTCCTCAGGGGATGAGCAGTCCCTGCTGGAACTTATCATCACCACCAAGAAGCGGGAGGCTCGCCAGATCCTGGACCAGA ATCATCTGCTTCACCATGTGCTGCATCTACCGCCCCCTCAAGCCCAGGACCAATAACCGCACAAGCCCCCGGGACAACACCCCTCTT A CAGCAGAAGCTACTTCAGGAAGCCTACGTGACCCCTAAGGACGATATCCGGCTGGTCGGGGAGCTGGTGACTGTCATTGGGGCTACTCCTTTGCACTCGTGCTGGGCTGCTACCTCATGTACTTCGCCCGAGGATTCCAGATGCTAGGCCCCTTCACCATCATGATTC GGACCGCTGGTTCCTGCGGGTGGAAGACAGGCAAGATCTCAACCGGCAGCGGATCCAACGCTACGCACAGGCCTTCCACACCCGGG GTGTCTCGAAGTACCTCCCGCAGCAGTGCCAATTGGGAAAGGCTTCGGCAAGGGACCCTGAGGAGAGACCTGCGTGGGATAATCAA  ${\tt CAGGGGTCTGGAGGAGGGGGAGAGCTGGGAATATCAGATCTGACTGCGTGTTCTCACTTCGCTTCCTGGAACTTGCTCTCATTTTC}$ AAGGAATGTACGTTGAGAATCACTGCTCCAGGCCTGCATTACTCCTTCAGCTCTGGGGCAGGGAAGCCCAAGCCACGCGCGCC TGGCAGGGCGTGAGGAACTCTCCTGTGGCCTGCTCATCACCCTTCCGACAGGAGCACTGCATGTCAGAGCACTTTAAAAACAGGCC  ${\tt GCAGAGCTTGTGGAAAGCGTGTGAGTGAGGGAGACAGGAACGGCTCTGGGGGTGGGAAGTGGGGCTAGGTCTTGCCAACTCCATCT}$ 

e.)

CACACATGGGGCCTCCCAGGAGTGCCCAGGACCTCGTGCTGTTGGCCTCTGAATCTATCGTCTCCAATCCGCTGTCCCACAGAAGC CATATAACCCACCTCTCTGTAAATGCCAGGAGCCATGGGGGAAACAGCGCTACACATAGCAGCCCTCTATGACAACCTGGAGGCCC CCATGGTGCTGATGGAGGCTGCCCCGGGGCTGGTCTTTGAGCCCATGACATCTGAGCTCTATGGAGGGTGAGGGCCCACGGGTCTG GGGTGAAGAGCAGGAGTGACGTCGGTAGTCAAGTCAGTCTCTGTGATGGATAATTTGGGAAAGACACAGGGGATCTGAGCCT CCTACTCTTTTTSTCTTCTCTCTCTCCCTTCCGTGTCAGTCCCTGACTGCCCATCACTTGAACGCCTGCCCCCTGAAATGCCAGGG GCCTAGAGAAGAGGAAGAGATGGGCAGCAGCTGGATCCCCTGGGAATCCTGAACACCCCGAGAGCTCCCTGTTCTCCATCCCAGGCT  $\tt CTGGGCCAGGTCAGACTGCACATCGCTGTTGTGAACCAGAACATGAACCTGGTGCGAGCCCTGCTTGCCCGCAGGGCCCAGT$ GTCTCTGCCAGAGCCACAGGCACTGCCTTCCGCCGTAGTCCCTGCAACCTCATCTACTTTGGGGAGCACCCTTTGTCCTTTGCTGC CTGTGTGAACAGTGAGGAGATCGTGCGCTGCTCATTGAGCATGGAGCTGACATCCGGGCCCAGGACTCCCTGGATGTACAACCTG TTGCTGTCCTACGACAGACATGGGGACCACCTGCAGCCCCTGGACCTCGTGCCCAATCACCAGGGTCTCACCCCTTTCAAGCTGGC  $\tt TGGAGTGGAGGGTAACACTGTGATGTTTCAGCACCTGATGCAGAAGCGGAAGCACCACGTGGACGTATGGACCACTGACCTCGA$ CTCTCTATGACCTCACAGAGATCGACTCCTCAGGGGATGAGCAGTCCCTGGTGGAACTTATCATCACCACCAAGAAGCGGGAGGCT  $\tt CGCCAGATCCTGGACCAGACCCCGGTGAAGGAGCTGGTGAGCCTCAAGTGGAAGGGGTACGGGCGGCCGTACTTCTGCATGCTGGG$ TGCCATATATCTGCTGTACATCATCTGCTTCACCATGTGCTGCATCTACCGCCCCCTCAAGCCCAGGACCAATAACCGCACGAGCC GTGACTGTCATTGGGGCTATCATCATCCTGCTGGTAGAGGTTCCAGACATCTTCAGAATGGGGGTCACTCGCTTCTTTGGACAGAC CATCCTTGGGGGCCCATTCCATGTCCTCATCACCTATGCCTTCATGGTGCTGGTGACCATGGTGATGCGGCTCATCAGTGCCA TTCAGCCTTCTATATCATCTTCCAGACAGAGGACCCCGAGGAGCTAGGCCACTTCTACGACTACCCCCATGGCCCTGTTCAGCACCT  ${\tt TCGAGCTGGTCCTTACCATCGATGGCCCAGCCAACTACAACGTGGACCTGCCCTTCATGTACAGCATCACCTATGCTGCCTTT$ GACGGGAGTATGGCCTGGGAGACCGCTGGTTCCTGCGGGTGGAAGACAGGCAAGATCTCAACCGGCAGCGGATCCAACGCTACGCA CAGGCCTTCCACACCCGGGGCTCTCAGGATTTGGACAAAGACTCAGTGGAAAAACTAGAGCTGGGCTGTCCCTTCAGCCCCCACCT GTCCCTTCCTATGCCCTCAGTGTCTCGAAGTACCTCCCGCAGCAGTGCCAATTGGGAAAGGCTTCGGCAAGGGACCCTGAGGAGAG ACCTGCGTGGGATAATCAACAGGGGTCTGGAGGACGGGGAGAGCTGGGAATATCAGATCTGACTGCGTGTTCTCACTTCGCTTCCT GGAACTTGCTCTCATTTTCCTGGGTGCATCAAACAAAACAAAAACCAAAACCCCAGAGGTCTCATCTCCCAGGCCCCCAGGGAAAA GAGGAGTAGCATGAACGCCAAGGAATGTACGTTGAGAATCACTGCTCCAGGCCTGCATTACTCCTTCAGCTCTGGGGCAGAGGAAG 

Fig. 8 / continuation 5

Figure 9:

A.

30 50 CGGGGCCCTGGGCTGCAGGAGGTTGCGGCCGCCGCCAGCATGGTGCCGGAGAAGG M V V P E K E 90 110 AGCAGAGCTGGATCCCCAAGATCTTCAAGAAGAAGACCTGCACGACGTTCATAGTTGACT Q S W I P K I F K K K T C T T F I V D S 130 150 170 CCACAGATCCGGGAGGGACCTTGTGCCAGTGTGGGCGCCCCGGACCGCCCACCCCGCAG T D P G G T L C Q C G R P R T A H P A V .210 TGGCCATGGAGGATGCCTTCGGGGCAGCCGTGGTGACCGTGTGGGACAGCGATGCACACA AMEDAF, GAAVVT VWD SDAH, T 250 270 290 TEKPTDAYGELDFTGAGRKH 310 330 350 ACAGCAATTTCCTCCGGCTCTCTGACCGAACGGATCCAGCTGCAGTTTATAGTCTGGTCA S N F L R L S D R T D P A A V Y S L V . T 390 CACGCACATGGGGGTTCCGTGCCCCGAACCTGGTGGTGTCAGTGCTGGGGGGATCGGGGG 430 450 510 AGAGCACAGGAGCCTGGATTGTCACTGGGGGTCTGCACACGGGCATCGGCCGCATGTTG S T G A W I V T G G L H T G I G R H V G 590 570 550 GTGTGGCTGTACGGGACCATCAGATGGCCAGCACTGGGGGCACCAAGGTGGTGGCCATGG V A V R D H Q M A S T G G T K V V A M G 630 GTGTGGCCCCCTGGGGTGTGGTCCGGAATAGAGACACCCTCATCAACCCCAAGGGCTCGT V A P W G V V R N R D T L I N P K G S F 690  ${\tt TCCCTGCGAGGTACCGGTGGCGGGTGACCCGGAGGACGGGGTCCAGTTTCCCCTGGACT}$ PARYRWRGDPEDGVQFPLDY 750 ACAACTACTCGGCCTTCTTCCTGGTGGACGACGCACACACGGCTGCCTGGGGGGGCGAGA NYSAFFLVDDGTHGCLGGEN 810 R F R L R L E S Y I S Q Q K T G V G G T 870 890 CTGGAATTGACATCCCTGTCCTGCTCCTCCTGATTGATGGTGATGAGAAGATGTTGACGC G I D I P V L L L L I D G D E K M L T R 910 930 950 I E N A T Q A Q L P C L L V A G S G G A 990 1010 CTGCGGACTGCCTGGCGGAGACCCTGGAAGACACTCTGGCCCCAGGGAGTGGGGGAGCCA A D C L A E T L E D T L A P G S G G A R 1050 1030 GGCAAGGCGAAGCCCGAGATCGAATCAGGCGTTTCTTTCCCAAAGGGGACCTTGAGGTCC

Fig. 9 / continue -- n 1

Q G E A R D R I R R F F P K G D L E V L 1110 1130 TGCAGGCCCAGGTGGAGAGGATTATGACCCGGAAGGAGCTCCTGACAGTCTATTCTTCTG Q A Q V E R I M T R K E L L T V Y S S E 1170 1150 AGGATGGGTCTGAGGAATTCGAGACCATAGTTTTGAAGGCCCTTGTGAAGGCCTGTGGGA D G S E E F E T I V L K A L V K A C G S 1230 GCTCGGAGGCCTCAGCCTACCTGGATGAGCTGCGTTTGGCTGTGGCTTGGAACCGCGTGG S E A S A Y L D E L R L A V A W N R V D 1290 1310  ${\tt ACATTGCCCAGAGTGAACTCTTTCGGGGGGGACATCCAATGGCGGTCCTTCCATCTCGAAG}$ I A Q S E L F R G D I Q W R S F H L E A 1330 1350 1370 CTTCCCTCATGGACGCCCTGAATGACCGGCCTGAGTTCGTGCGCTTGCTCATTTCCC SLMDALLNDRPEFVRLLISH 1410 1430 ACGCCTCAGCCTGGCCCACTTCCTGACCCCGATGCGCCCTGGCCCAACTCTACAGCGCGG G L S L G H F L T P N R L A Q L Y S A A 1450 1470 1490  $\tt CGCCCTCCAACTCGCTCATCCGCAACCTTTTGGACCAGGCGTCCCACAGCGCAGGCACCCA$ PSNSLIRNLLDQASHSAGTK 1530 1550  ${\tt AAGCCCCAGCCCTAAAAGGGGGAGCTGCGGAGCTCCGGCCCCCTGACGTGGGGCATGTGC}$ APALKGGAAELRPPDVGHVL 1590 1610 TGAGGATGCTGCGGGAAGATGTGCGCGCCGAGGTACCCCTCCGGGGCGCCCTGGGACC RMLLGKMCAPRYPSGGAWDP 1650 1670 1630  $\tt CTCACCCAGGCCAGGGCTTCGGGGAGAGCATGTATCTGCTCTCGGACAAGGCCACCTCGC$ HPGQGFGESMYLLSDKATSP 1690 1710 1730  ${\tt CGCTCTCGCTGGATGCTGGCCTCGGGCAGCCCCCTGGAGCGACCTGCTTCTTTGGGCAC}$ L S L D A G L G Q A P W S D L L L W A L 1790 1770  ${\tt TGTTGCTGAACAGGGCACAGATGGCCATGTACTTCTGGGAGATGGGTTCCAATGCAGTTT}$ L L N R A Q M A M Y F W E M G S N A V S 1830 1850  ${\tt CCTCAGCTCTTGGGGGCCTGTTTGCTGCTCCGGGTGATGGCACGCCTGGAGCCTGACGCTG}$ S A L G A C L L R V M A R L E F D A E 1870 1890 1910 AGGAGGCAGCACGGAGAAGACCTGGCGTTCAAGTTTGAGGGGATGGGCGTTGACCTCT E A A R R K D L A F K F E G M G V D L F 1950 G E C Y R S S E V R A A R L L R R C P 1990 2010 2030  $\tt CGCTCTGGGGGGATGCCACTTGCCTCCAGCTGGCCATGCAAGCTGACGCCCGTGCCTTCT$ L W G D A T C L Q L A M Q A D A R A F F 2070 2090 TTGCCCAGGATGGGGTACAGTCTCTGCTGACACAGAAGTGGTGGGGAGATATGGCCAGCA A Q D G V Q S L L T Q K W W G D M A S T 2150 2110 2130 CTACACCCATCTGGGCCCTGGTTCTCGCCTTCTTTTGCCCTCCACTCATCTACACCCGCC T P I W A L V L A F F C P P L I Y T R L 2170 2190 2210 TCATCACCTTCAGGAAATCAGAAGAGGAGCCCACACGGGAGGAGCTAGAGTTTGACATGG I T F R K S E E P T R E E L E F D M D 2230 2250 ATAGTGTCATTAATGGGGAAGGGCCTGTCGGGACGGGGACCCCAGCCGAGAAGACGCCGC SVINGEGPVGTADPAEKTPL 2310 2290

Fig. 9 / continue ≒on 2

TGGGGGTCCCGCCCCAGTCGGGCCCTCCGGGTTGCTGCGGGGGCCCGCTGCGGGGGCCCCC G V P R Q S G R P G C C G G R C G G R R 2390 2370 GGTGCCTACGCCGCTGGTTCCACTTCTGGGGCGTGCCGGTGACCATCTTCATGGGCAACG CLRRWFHFWGVPVTIFMGNV 2410 2430 2450 TGGTCAGCTACCTGCTGTTCCTGCTGCTTTTCTCGCGGGTGCTGCTCGTGGATTTCCAGC V S Y L L F L L L F S R V L L V D F Q P 2510 2470 2490  $\tt CGGCGCCGGCCCGGGTCCTGGAGCTGCTGTTCTATTTCTGGGCTTTCACGCTGCTGTGCG$ APPGSLELLYFWAFTLLCE 2550 2570 2530 AGGAACTGCGCCAGGGCCTGAGCGGAGGCGGGGGCAGCCTCGCCAGCGGGGGCCCCGGGC E L R Q G L S G G G G S L A S G G P G P 2610 2630  $\tt CTGGCCATGCCTCACTGAGCCAGCGCCTGCGCCTCTACCTCGCCGACAGCTGGAACCAGT$  $\label{eq:constraints} \textbf{G} \quad \textbf{H} \quad \textbf{A} \quad \textbf{S} \quad \textbf{L} \quad \textbf{S} \quad \textbf{Q} \quad \textbf{R} \quad \textbf{L} \quad \textbf{R} \quad \textbf{L} \quad \textbf{Y} \quad \textbf{L} \quad \textbf{A} \quad \textbf{D} \quad \textbf{S} \quad \textbf{W} \quad \textbf{N} \quad \textbf{Q} \quad \textbf{C}$ 2650 2670 2690 GCGACCTAGTGGCTCTCACCTGCTTCCTCCTGGGCGTGGGCTGCCGGCTGACCCCGGGTT D L V A L T C F L L G V G C R L T P G L 2750 2730 TGTACCACCTGGGCCGCACTGTCCTCTGCATCGACTTCATGGTTTTCACGGTGCGGCTGC Y H L G R T V L C I D F M V F T V R L L 2770 2790 . 2910 TTCACATCTTCACGGTCAACAACAGCTGGGGCCCAAGATCGTCATCGTGAGCAAGATGA HIFTVNKQLGPKIVIVSKMM 2830 2850 2870 TGAAGGACGTGTTCTTCTTCTTCTTCCTCGGCGTGTGGCTGGTAGCCTATGGCGTGG 2910 CCACGGGGGCTCCTGAGGCCACGGGACAGTGACTTCCCAAGTATCCTGCGCCGCGTCT TEGLLRPRDSDFPSILRRVF 2970 2990 TCTACOGTCCCTACCTGCAGATCTTCGGGCAGATTCCCCAGGAGGACATGGACGTGGCCC Y R P Y L Q I F G Q I P Q E D M D V A L 3010 3030 3050 TCATGGAGCACAGCAACTGCTCGGAGCCCGGCTTCTGGGCACACCCTCCTGGGGCCC MEHSNCSSEPGFWAHPPGAQ 3090 AGGCGGCACCTGCCTCTCCCACTATGCCAACTGGCTGGTGGTGCTCCTCCTCATCT AGTCVSQYANWLVVLLLVIF 3130 3150 3170 TCCTGCTCGTGGCCAACATCCTGCTGGTCAACTTGCTCATTGCCATGTTCAGTTACACAT LLVANILLVNLLIAMFSYTF 3190 3210 3230 TCGGCAAAGTACAGGGCAACAGCGATCTCTACTGGAAGGCGCAGCGTTACCGCCTCATCC 3290 3270 GGGAATTCCACTCTCGGCCCGCGCTGGCCCCGCCCTTTATCGTCATCTCCCACTTGCGCC E F H S R P A L A P P F I V I S H L R L 3330 TCCTGCTCAGGCAATTGTGCAGGCGACCCCGGAGCCCCAGCCGTCCTCCCCGGCCCTCG L L R Q L C R R P R S P Q P S S P A L E 3410 3390  ${\tt AGCATTTCCGGGTTTACCTTTCTAAGGAAGCCGAGCGGAAGCTGCTAACGTGGGAATCGG}$ H F R V Y L S K E A E R K L L T W E S V 3470 3450 TGCATAAGGAGAACTTTCTGCTGGCACGCGCTAGGGACAAGCGGAGAGCGACTCCGAGC H K E N F L L A R A R D K R E S D S E R 3530 3510  $\tt GTCTGAAGCGCACGTCCCAGAAGGTGGACTTGGCACTGAAACAGCTGGGACACATCCGCG$ L K R T S Q K V D L A L K Q L G H I R E

Fig. 9 / continue n 3 AGTACGAACAGCGCCTGAAAGTGCTGGAGCGGGAGGTCCAGCAGTGTAGCCGCGTCCTGG Y E Q R L K V L E R E V Q Q C S R V L G GGTGGGTGGCCGAGGCCCTGAGCCGCTCTGCCTGCCCCCAGGTGGGCCGCCACCCC WVAEALSRSALLPPGGPPPP CTGACCTGCCTGGGTCCAAAGACTGAGCCCTGCTGGCGGACTTCAAGGAGAAGCCCCCAC DLPGSKD \* AGGGGATTTTGCTCCTAGAGTAAGGCTCATCTGGGCCTCGGCCCCCGCACCTGGTGGCCT TGTCCTTGAGGTGAGCCCCATGTCCATCTGGGCCACTGTCAGGACCACCTTTGGGAGTGT CATCCTTACAAACCACAGCATGCCCGGCTCCTCCCAGAACCAGTCCCAGCCTGGGAGGAT CAAGGCCTGGATCCCGGGCCGTTATCCATCTGGAGGCTGCAGGGTCCTTGGGGTAACAGG GACCACAGACCCCTCACCACTCACAGATTCCTCACACTGGGGAAATAAAGCCATTTCAGA **GGAAAAAAAAAAAAAAA** 

MVVPEKEQSWIPKIFKKKTCTTFIVDSTDPGGTLCQCGRPRTAHPAVAMEDAFGAAVVTWDSDAHTTEKPTDAYELDFTGAG
SNFLRLSDATDPAAVYSLVTRTWGFRAPNLVVSVLGGSGGPVLQTWLQDLLRRGLVRAAQSTGAWIVTGGLHTGIGRHVGVAV
QMASTGGTKVVAMGVAPWGVVRNRDTLINPKGSFPARYRWRGDPEDGVQFPLDYNYSAFFLVDDGTHGCLGGENRFRLRLESY
QKTGVGGTGIDPVLLLLIDGDEKNLTRIENATQAHVPCLLVAGSRGLGMPGGTLEAHLAQDGDHKANQSTNQLLLPRDLSLK
SIDRKTLQSYSERLAVAWNRVDLAQSELFRGDIQWRSFHLEASLMDALLNDRPEFVRLLISHGLSLGHFLTPWRLAQLYSAAE
LIRNLLDQASHSAGTKAPALKGGAAELRPPDVGHVLRMLLGKMCAPRYPSGGAWDPHPGQGFGESMYLLSDKATSPLSLDAGI
PWSDLLLWALLLNRAQMAMYFWEMGSNAVSSALGACLLLRVMARLEPDAEEAARRKDLAFKFEGMGVDLFGECYRSSEVRAAF
RRCPLWGDATCLQLAMQADARAFFAQDGVQSLLTQKWWGDDMASTTPIWALVLAFFFPPLIYTRLITFRKSEEEPTREELEFD
INGEGPVGTADPAEKTPLGVPRQSGRPGCCGGRCGGRRCLRRWFHFWGVPVTIFMCNVVSYLLFLLFSRVLLVDFQPAPPGS
LLYFWAFTLLCEELRQGLSGGGGSLASGGPGPGHASLSQRLRLYLADSWNQCDLVALTCFLLGVGCRLTPGLYHLGRTVLCIF
FTVRLHIFTVNKQLGFKIVIVSKMMKDVFFFLFFLGVWLVAYGVATEGLLRPRDSDFPSILRRVFYRPYLQIFGQIPQDDMI
MEHSNCSSEPGFWHPPGAQAGTCVSQYANWLVVLLLVIFLLVANILLVNLLIAMFSYTFGKVQGNSDLYWKAQRYRLIREFF
ALAPPFIVISHLRLLLRQLCRRPRSPQPSSPALEHFRVYLSKEAERKLLTWESVHKENFLLARARDKRESDSERLKRTSQKVI
KQLGHIREYEQRLKVLEREVQQCSRVLGWVAEALSRSALLPPGGPPPPDLPGSKD

в.)

ATCCAATGCCGTCCTTCCATCTCGAAGCTTCCCTCATGGACGCCCTGCTGAATGACCGG CCTGAGTTCGTGCGCTTGCTCATTTCCCACGGCCTCAGCCTGGGCCACTTCCTGACCCCG ATGCGCCTGGCCCAACTCTACAGCGCGCGCCCCTCCAACTCGCTCATCCGCAACCTTTTG GACCAGGCGTCCCACAGCGCAGGCACCAAAGCCCCAGCCCTAAAAGGGGGAGCTGCGGAG CTCCGGCCCCTGACGTGGGGCATGTGCTGAGGATGCTGCTGGGGAAGATGTGCGCGCCG AGATGTATCTGCTCTCGGACAAGGCCACCTCGCCGCTCTCGCTGGATGCTGGCCTCGGGC MYLLSDKATSPLSLDAGLGQ AGGCCCCCTGGAGCGACCTGCTTCTTTGGGCACTGTTGCTGAACAGGGCACAGATGGCCA APWSDLLLWALLLNRAQMAM TGTACTTCTGGGAGATGGGTTCCAATGCAGTTTCCTCAGCTCTTGGGGCCTGTTTGCTGC Y F W E M G S N A V S S A L G A C L L L Fig. 9 / continue 4 4

```
510 .
     490
TCCGGGTGATGGCACGCCTGGAGCCTGACGCTGAGGAGGCAGCACGGAGGAAAGACCTGG
 RVMARLEPDAEEAARRKDLA
                    570
                                    590
CGTTCAAGTTTGAGGGGATGGGCGTTGACCTCTTTGGCGAGTGCTATCGCAGCAGTGAGG
 F K F E G M G V D L F G E C Y R S S E V
                                    650
                    630
TGAGGGCTGCCGCCTCCTCCGTCGCTGCCCGCTCTGGGGGGATGCCACTTGCCTCC
 RAARLLLRRCPLWGDATCLQ
                    690
AGCTGGCCATGCAAGCTGACGCCCGTGCCTTCTTTGCCCAGGATGGGGTACAGTCTCTGC
 L A M Q A D A R A F F A Q D G V Q S L L
                    750
                                    770
TGACACAGAAGTGGTGGGGAGATATGGCCAGCACTACACCCATCTGGGCCCTGGTTCTCG
 T Q K W W G D M A S T T P I W A L V L A
                    810
{\tt CCTTCTTTTGCCCTCCACTCATCTACACCCGCCTCATCACCTTCAGGAAATCAGAAGAGG}
 F F C P P L I Y T R L I T F R K S E E E
     850
                    870
AGCCCACACGGGAGGAGCTAGAGTTTGACATGGATAGTGTCATTAATGGGGAAGGGCCTG
 PTREELEFDMDSVINGEGPV
                    930
                                    950
TCGGGACGGCGGACCCAGCCGAGAAGACGCCGCTGGGGGTCCCGCGCCAGTCGGGCCGTC
 G T A D P A E K T P L G V P R Q S G R P
                    990
                                  1010
G C C G G R C G G R R C L R R W F H F W
                                  1070
                   1050
EGGGCGTGCCGGTGACCATCTTCATGGGCAACGTGGTCAGCTACCTGCTGTTCCTGCTGC
 G V P V T I F M G N V V S Y L L F L L
                                  1130
    1090
                   1110
TTTTCTCGCGGGTGCTGCTCGTGGATTTCCAGCCGGCCCCCGGCTCCCTGGAGCTGC
 FSRVLLVDFQPAPPGSLELL
                    1170
                                   1190
    1150
TGCTCTATTTCTGGGCTTCACGCTGCTGTGCGAGGAACTGCGCCAGGGCCTGAGCGGAG
  LYFWAFTLLCEELRQGLSGG
                                   1250
                    1230
GCGGGGGCAGCCTCGCCAGCGGGGCCCGGGCCTGGCCATGCCTCACTGAGCCAGCGCC
  G G S L A S G G P G P G H A S L S Q R L
    1270
                    1290
TGCGCCTCTACCTCGCCGACAGCTGGAACCAGTGCGACCTAGTGGCTCTCACCTGCTTCC
  RLYLADSWNQCDLVALTCFL
                                   1370
    1330
                    1350
TCCTGGGCGTGGCCTGCCGGCTGACCCCGGGTTTGTACCACCTGGGCCGCACTGTCCTCT
  LGVGCRLTPGLYHLGRTVLC
                    1410
                                   1430
{\tt GCATCGACTTCATGGTTTTCACGGTGCGGCTGCTTCACATCTTCACGGTCAACAACAGC}
  I D F M V F T V R L L H I F T V N K Q L
                    1470
                                   1490
     1450
TGGGGCCCAAGATCGTCATCGTGAGCAAGATGATGAAGGACGTGTTCTTCTTCCTCTTCT
  G P K I V I V S K M M K D V F F F L F F
                    1530
                                   1550
TCCTCGGCGTGTCGCTAGCCTATGGCGTGGCCACGGAGGGGCTCCTGAGGCCACGGG
  LGVWLVAYGVATEGLLRPRD
                    1590
ACAGTGACTTCCCAAGTATCCTGCGCCGCGTCTTCTACCGTCCCTACCTGCAGATCTTCG
  S D F P S I L R R V F Y R P Y L Q I F G
                    1650
                                   1670
GGCAGATTCCCCAGGAGGACATGGACGTGGCCCTCATGGAGCACAGCAACTGCTCGTCGG
  Q I P Q B D M D V A L M E H S N C S S E
                                   1730
                    1710
     1690
AGCCCGGCTTCTGGGCACACCCTCCTGGGGCCCAGGCGGGCACCTGCGTCTCCCAGTATG
```

Fig. 9 / continuation 5 PGFWAHPPGAQAGTCVSQYA 1790 1770 1750 CCAACTGGCTGGTGGTGCTCCTCGTCATCTTCCTGCTCGTGGCCAACATCCTGCTGG N W L V V L L L V I F L L V A N I L L V 1850 1830 TCAACTTGCTCATTGCCATGTTCAGTTACACATTCGGCAAAGTACAGGGCAACAGCGATC N L L I A M F S Y T F G K V Q G N S D L 1910 1890 1870 TCTACTGGAAGGCGCAGCGTTACCGCCTCATCCGGGAATTCCACTCTCGGCCCGCGCTGG W K A Q R Y R L I R E F H S R P A L A 1950 1970  $\tt CCCCGCCCTTTATCGTCATCTCCCACTTGCGCCTCCTGCTCAGGCAATTGTGCAGGCGAC$ P P F I V I S H L R L L R Q L C R R P 2010 2030 1990  $\tt CCCGGAGCCCCAGCCGTCCTCCCCGGCCCTCGAGCATTTCCGGGTTTACCTTTCTAAGG$ RSPQPSSPALEHFRVYLSKE 2090 2050 2070 AAGCCGAGCGGAAGCTGCTAACGTGGGAATCGGTGCATAAGGAGAACTTTCTGCTGGCAC AERKLLTWESVHKENFLLA.R 2110 2130 2150 GCGCTAGGGACAAGCGGGAGAGCGACTCCGAGCGTCTGAAGCGCACGTCCCAGAAGGTGG ARDKRESDSERLKRTSQKVD 2210 2190 ACTTGGCACTGAAACAGCTGGGACACATCCGCGAGTACGAACAGCGCCTGAAAGTGCTGG LALKQLGHIREYEQRLKVLE 2270 2250 REVQQCSRVLGWVAEALSRS 2310 2330 2290 A L L P P G G P P P P D L P G S K D \* 2370 2390 -2350 CCCTGCTGGCGGACTTCAAGGAGAAGCCCCCACAGGGGATTTTGCTCCTAGAGTAAGGCT 2430 2450 2410 CATCTGGGCCTCGGCCCCGCACCTGGTGGCCTTGTCCTTGAGGTGAGCCCCATGTCCAT 2510 2470 2490  $\tt CTGGGCCACTGTCAGGACCACCTTTGGGAGTGTCATCCTTACAAACCACAGCATGCCCGG$ 2550 2570 CTCCTCCCAGAACCAGTCCCAGCCTGGGAGGATCAAGGCCTGGATCCCGGGCCGTTATCC 2610 2630 ATCTGGAGGCTGCAGGGTCCTTGGGGTAACAGGGACCACAGACCCCTCACCACTCACAGA 2670 2690 

MYLLSDKATSPLSLDAGLGQAPWSDLLLWALLLNRAQMAMYFWEMGSNAVSSALGACLLLRVMARLEPDAEEAARRKDLAFKFEGM
GVDLFGECYRSSEVRAARLLLRRCPLWGDATCLQLAMQADARAFFAQDGVQSLLTQKWWGDMASTTPIWALVLAFFCPPLIYTRLI
TFRKSEEEPTREELEFDMDSVINGEGPVGTADPAEKTPLGVPRQSGRPGCCGGRCGGRRCLRRWFHFWGVPVTIFNGNVVSYLLFL
LLFSRVLLVDFQPAPPGSLELLLYFWAFTLLCEELRQGLSGGGGSLASGGPGPGHASLSQRLRLYLADSWNQCDLVALTCFLLGVG
CRLTPGLYHLGRTVLCIDFMVFTVRLLHIFTVNKQLGPKIVIVSKMMKDVFFFLFFLGVWLVAYGVATEGLLRPRDSDFPSILRRV
FYRPYLQIFGQIPQEDMDVALMBHSNCSSEPGFWAHPPGAQAGTCVSQYANWLVVLLLVIFLLVANILLVNLLIAMFSYTFGKVQG
NSDLYWKAQRYRLIREFHSRPALAPPFIVISHLRLLLRQLCRRFRSPQPSSPALEHFRVYLSKEAERKLLTWESVHKENFLLARAR
DKRESDSERLKRTSQKVDLALKQLGHIREYEQRLKVLEREVQQCSRVLGWVAEALSRSALLPPGGPPPPDLPGSKD

Fig. 10

10 ATTAAAGTTTATAAAACAGTGGCTGGATGGTTCGAGGATGCACGTGGACAGAAGACGTCG M V G G C R W T E D V E 90 110 AGCCTGCAGAAGTAAAGGAAAAGATGTCCTTTCGGGCAGCCAGGCTCAGCATGAGGAACA PAEVKEKMSFRAARLSMRNR 130 150 170 GAAGGAATGACACTCTGGACAGCACCCGGACCCTGTACTCCAGCGCGTCTCGGAGCACAG RNDTLDSTRTLYSSASRSTD 210 230 ACTTGTCTTACAGTGAAAGCGCCAGCTTCTACGCTGCCTTCAGGACACAGACGTGCCCAA L S Y S E S A S F Y A A F R T Q T C P I 270 290 TCATGGCTTCTTGGGGACTTGGTGAATTTTATTCAAGCAAATTTTAAGAAACGAGAATGTG M A S W D L V N F I Q A N F K K R E C V . 330 310 TCTTCTTTACCAAAGATTCCAAGGCCACGGAGAATGTGTGCAAGTGTGGCTATGCCCAGA F F T K D S K A T E N V C K C G Y A Q S 390 GCCAGCACATGGAAGGCACCCAGATCAACCAAAGTGAGAAATGGAACTACAAGAAACACA Q H M E G T Q I N Q S E K W N Y K K H T 470 430 450 CCAAGGAATTTCCTACCGACGCCTTTGGGGATATTCAGTTTGAGACACTGGGGAAGAAAG KEFPTDAFGDIQFETLGKKG 490 510 530 GGAAGTATATACGTCTGTCCTGCGACACGGACGCGGAAATCCTTTACGAGCTGCTGACCC KYIRLSCDTDAEILYELLTQ 570 590 AGCACTGGCACCTGAAAACACCCAACCTGGTCATTTCTGTGACCGGGGGCGCCAAGAACT H W H L K T P N L V I S V T G G A K N F 630 TCGCCCTGAAGCCGCGCATGCGCAAGATCTTCAGCCGGCTCATCTACATCGCGCAGTCCA A L K P R M R K I F S R L I Y I A Q S K 690 AAGGTGCTTGGATTCTCACGGGAGGCACCCATTATGGCCTGATGAAGTACATCGGGGAGG G A W I L T G G T H Y G L M K Y I G E V 730 750 TGGTGAGAGATAACACCATCAGCAGGAGTTCAGAGGAGAATATTGTGGCCATTGGCATAG V R D N T I S R S S E E N I V A I G I A 810 830 CAGCTTGGGGCATGGTCTCCAACCGGGACACCCTCATCAGGAATTGCGATGCTGAGGGCT A W G M V S N R D T L I R N C D A E G Y 870 ATTTTTTAGCCCAGTACCTTATGGATGACTTCACAAGAGATCCACTGTATATCCTGGACA F L A Q Y L M D D F T R D P L Y I L D N 950 910 930 ACAACCACACACTTTGCTGCTCGTGGACAATGGCTGTCATGGACATCCCACTGTCGAAG N H T H L L L V D N G C H G H P T V E A 990 1010 CAAAGCTCCGGAATCAGCTAGAGAAGTATATCTCTGAGCGCACTATTCAAGATTCCAACT K L R N Q L E K Y I S E R T I Q D S N Y 1030 1050 ATGGTGGCAAGATCCCCATTGTGTGTTTTGCCCAAGGAGGTGGAAAAGAGACTTTGAAAG G G K I P I V C F A Q G G G K E T L K A 1110 1130  ${\tt CCATCANTACCTCCATCANANATAAAATTCCTTGTGTGGTGGTAGGCTCGGGCCAGA}$ INTSIKNKIPCVVVEGSGQI 1190 1170 TCGCTGATGTGATCGCTAGCCTGGTGGAGGTGGACGATGCCCTGACATCTTCTGCCGTCA A D V I A S L V E V E D A L T S S A V K 1230 1210

PCT/EP01/08309

Fig. 10 / continuation 1

AGGAGAGCTGCTGCGCTTTTTACCCCGCACGTGTCCCGGCTGCCTGAGGAGGAGACTG EKLVRFLPRTVSRLPEETE 1270 1290 AGAGTTGGATCAAATGGCTCAAAGAAATTCTCGAATGTTCTCACCTATTAACAGTTATTA S W I K W L K E I L B C S H L L T V I K 1350 AAATGGAAGAAGCTGGGGATGAAATTGTGAGCAATGCCATCTCCTACGCTCTATACAAAG MEEAGDEIVSNAISYALYKA 1410 1430 CCTTCAGCACCAGTGAGCAAGACAAGGATAACTGGAATGGGCAGCTGAAGCTTCTGCTGG 1490 1470 AGTGGAACCAGCTGGACTTAGCCAATGATGAGATTTTCACCAATGACCGCCGATGGGAGA W N Q L D L A N D E I F T N D R R W E K 1530 1550 AGAGCAAACCGAGGCTCAGAGACACAATAATCCAGGTCACATGGCTGGAAAATGGTAGAA SKPRLRDTIIQVTWLENGRI 1590 TCAAGGTTGAGAGCAAAGATGTGACTGACGGCAAAGCCTCTTCTCATATGCTGGTGGTTC K V E S K D V T D G K A S S H M L V V L 1630 1650 1670 TCAAGTCTGCTGACCTTCAAGAAGTCATGTTTACGGCTCTCATAAAGGACAGACCCAAGT K S A D L Q E V M F T A L I K D R P K F 1690 1710 · 1730 TTGTCCGCCTCTTTCTGGAGAATGGCTTGAACCTACGGAAGTTTCTCACCCATGATGTCC 1770 1750 TCACTGAACTCTTCTCCAACCACTTCAGCACGCTTGTGTACCGGAATCTGCAGATCGCCA TELFSNHFSTLVYRNLQIAK 1850 1830 1810 AGAATTCCTATAATGATGCCCTCCTCACGTTTGTCTGGAAACTGGTTGCGAACTTCCGAA NSYNDALLTFVWKLVANFRR 1910 1890 GAGGCTTCCGGAAGGAAGACAGAAATGGCCGGGACGAGATGGACATAGAACTCCACGACG G F R K E D R N G R D E M D I E L H D V 1970 1950 TGTCTCCTATTACTCGGCACCCCCTGCAAGCTCTCTTCATCTGGGCCATTCTTCAGAATA S PIT T R H P L Q A L F I W A I L Q N K 2030 1990 2010 AGAAGGAACTCTCCAAAGTCATTTGGGAGCAGACCAGGGGCTGCACTCTGGCAGCCCTGG 2070 GAGCCAGCAAGCTTCTGAAGACTCTGGCCAAAGTGAAGAACGACATCAATGCTGCTGGGG ASKLLKTLAKVKN DINAAGE 2150 2110 2130 AGTCCGAGGAGCTGGCTAATGAGTACGAGACCCGGGCTGTTGGTGAGTCCACAGTGTGGA SEELANEYETRAVGESTVWN 2210 2190 ATGCTGTGGTGGGCGCGGATCTGCCATGTGGCACAGACATTGCCAGCGGCACTCATAGAC A V V G A D L P C G T D I A S G T H R P 2230 2250 CAGATGGTGGAGAGCTGTTCACTGAGTGTTACAGCAGCGATGAAGACTTGGCAGAACAGC D G G E L F T E C Y S S D E D L A E Q L 2330 2310 TGCTGGTCTATTCCTGTGAAGCTTGGGGTGGAAGCAACTGTCTGGAGCTGGCGGTGGAGG LVYSCEAWGGSNCLELAVEA 2390 2350 2370  ${\tt CCACAGACCAGCATTTCATCGCCCAGCCTGGGGTCCAGAATTTTCTTAAGCAATGGT}$ T D Q H F I A Q P G V Q N F L S K Q W Y 2450 2430 2410 ATGGAGAGATTTCCCGAGACACCAAGAACTGGAAGATTATCCTGTGTCTGTTTATTATAC GEISRDTKNWKIILCLFIIP

2510 2490 CCTTGGTGGGCTGTGGCTTTGTATCATTTAGGAAGAACCTGTCGACAAGCACAAGAAGC L V G C G F V S F R K K P V D K H K K L 2550 2570 TGCTTTGGTACTATGTGGCGTTCTTCACCTCCCCCTTCGTGGTCTTCTCCTGGAATGTGG LWYYVAFFTSPFVVFSWNVV 2630 2590 2610  ${\tt TCTTCTACATCGCCTTCCTCCTGCTGTTTGCCTACGTGCTCATGGATTTCCATTCGG}$ FYIAFLLLFAYVLL M D F H S V 2670 2690 TGCCACACCCCCCGAGCTGGTCCTGTACTCGCTGGTCTTTGTCCTCTTCTGTGATGAAG PHPPELVLYSLVFVLFCDEV 2710 2730 TGAGACAGGGCCGGCCGGCTGCTCCCAGTGCGGGGCCCGCCAAGCCCACGCCCACCCGGA RQGRPAAPSAGPAKPTPTRN 2790 2810 ACTCCATCTGGCCCGCAAGCTCCACACGCAGCCCCGGTTCCCGCTCACGCCACTCCTTCC SIW PASSTRSPGSRSR HSF H - 2850 2870 ACACTTCCCTGCAAGCTGAGGGTGCCAGCTCTGGCCTTGGCCAGCCCAGAAAGGGGTGGA T S L Q A E G A S S G L G Q P R K G W T 2910 2930 CATTTAAAAATCTGGAAATGGTTGATATTTCCAAGCTGCTGATGTCCCTCTCTGTCCCTT F K N L B M V D I S K L L M S L S V P F 2970 TCTGTACGCAGTGGTACGTAAATGGGGTGAATTATTTTACTGACCTGTGGAATGTGATGG C T Q W Y V N G V N Y F T D L W N V M D 3030 3050 3010 ACACGCTGGGGCTTTTTTACTTCATAGCAGGAATTGTATTTCGGCAAGGGATCCTTAGGC TLGLFYFIAGIVFRQGILRQ 3070 3090 3110 AGAATGAGCAGCGCTGGAGGTGGATATTCCGTTCGGTCATCTACGAGCCCTACCTGGCCA NEQRWRWIFRSVIYEPYLAM 3150 3170 TGTTCGGCCAGGTGCCCAGTGACGTGGATGGTACCACGTATGACTTTGCCCACTGCACCT F G Q V P S D V D G T T Y D F A H C T F 3190 3210 3230 TCACTGGGAATGAGTCCAAGCCACTGTGTGTGGAGCTGGATGAGCACAACCTGCCCCGGT TGNESKPLCVELDEHNLPRF 3290 3270 TCCCCGAGTGGATCACCATCCCCCTGGTGTGCATCTACATGTTATCCACCAACATCCTGC PEWITIPLVCIYMLSTNILL 3350 3310 3330 TGGTCAACCTGCTGGTCGCCATGTTTGGCTACACGGTGGGCACCGTCCAGGAGAACAATG V N L L V A M F G Y T V G T V Q E N N D 3390 ACCAGGTCTGGAAGTTCCAGAGGTACTTCCTGGTGCAGGAGTACTGCAGCCGCCTCAATA V W K F Q R Y F L V Q E Y C S R L N I 3470 . 3450 TCCCCTTCCCCTTCATCGTCTTCGCTTACTTCTACATGGTGGTGAAGAAGTGCTTCAAGT PFPFIVFAYFYMVVKKCFKC 3510 3530 GTTGCTGCAAGGAGAAAACATGGAGTCTTCTGTCTGCTGTGAGTGGTTTATCCATGTGT C C K E K N M E S S V C C E W F I H V Y 3590 3550 3570 ACTTGGGATCAGAAGCAGCGATTAATTTCAGGGAAGGATGCCTGCATCCAGTGATTGGAA LGSEAAINFREGCLHPVIGS 3630 3650 3610 GCTGGACCCCAGGCTGGCTGGTCTGGACATCCACACGCATTCTCACATGCAGTGCCGGCT W T P G W L V W T S T R I L T C S A G W 3690 GGCCAGCAGCAGGGAGTCTCAGTGTCACCACACATAGCAGCTGGGTTCCTGCAAAAAGCA

Fig. 10 / continuation 3

MVGGCRWTEDVEPAEVKEKMSFRAARLSMRNRRNDTLDSTRTLYSSASRSTDLSYSESASFYAAFRTQTCPIMASWDLVNFIQANF
KKRECVFFTKDSKATENVCKCGJAQSQHMEGTQINQSEKWNYKKHTKEFFTDAFGDIQFETLGKKGKYIRLSCDTDAEILYELLTQ
HWHLKTENLVISVTGGAKNFALKPRMRKIFSRLIYIAQSKGAWILTGGTHYGLMKYIGEVVRDNTISRSSEENIVAIGIAAWGMVS
NRDTLIRNCDAEGYFLAQYLMDDFTRDPLYILDNNHTHLLLVDNGCHGHPTVEAKLRNQLBKYISERTIQDSNYGGKIPIVCFAQG
GGKETLKAINTSIKNKIPCVVVEGSGQIADVIASLVEVEDALTSSAVKEKLVRFLPRTVSRLPBEETESWIKWLKEILECSHLLTV
IKMEEAGDEIVSNAISYALYKAFSTSSEQDKDNWNGQLKLLLEWNCLDLANDEIFTNDRRWBKSKPRLRDTIIQVTWLENGRIKVES,
KDVTDGKASSHMLVVLKSADLQEVMFTALIKDRPKFVRLFLENGLNLKKFLTHDVLITELFSNHFSTLVYRNLQIAKNSYNDALLTF
VWKLVANFRRGFRKEDRNGRDEMDIELHDVSPITRHPLQALFIWAILQNKKELSKVIWEQTRGCTLAALGASKLLKTLAKVKNDIN
AAGESEELANEYETRAVGESTVWNAVVGADLPCGTDIASGTHRPDGGELFTECYSSDEDLAEQLLVYSCEAWGGSNCLELAVEATD
QHFIAQPGVQNFLSKQWYGEISRDTKNWKIILCLFIIPLVGCGFVSFRKKFVDKHKKLLMYYVAFFTSPFVVFSWNVVYTAFLLL
FAYVLIMDFHSVPHPPELVLYSLVFVLFCDEVRQGRPAAPSAGPAKPTPTRNSIWPASSTRSPGSRSRHSFHTSLQAEGASGGGQ
PRKGWTFKNLEMVDISKLLMSLSVPFCTQMYVNGVNYFTDLMNVMDTLGLPYFIAGIVFRQGILRQNEQRWRWIFRSVIYEPYLAM
FGQVPSDVDGTTYDFAHCTFTGNESKPLCVELDEHNLPRFPEWITIPLVCIYMLSTNILLVNLLVAMFGYTVGTVQENNDQVWKFQ
RYFLVQEYCSRLNIIFFFFTVFAYFYMVVKKCFKCCCKEKNMESSVCCEWFIHVYIGSEAAINFREGCLHEVIGSWTFGWLVWTSTR
ILTCSAGWPAAGSLSVTTHSSWVPAKSSKSQAHPDRTGRECDSASGWEGQPARWVEESVALFGHRGPVWPPTTLGITELNAFVL

в.

2290 2310  ${\tt TGCTGGTCTATTCCTGTGAGCTTGGGGTGGAAGCAACTGTCTGGAGCTGGCGGTGGAGG}$ LVYSCEAWGGSNCLELAVEA 2350 2370 2390 T D Q H F I A Q P G V Q N F L S K Q W Y 2410 2430 2450 ATGGAGAGATTTCCCGAGACACCAAGAACTGGAAGATTATCCTGTGTCTGTTTATTATAC G E I S R D T K N W K I I L C L F I I P 2470 2490 2510  ${\tt CCTTGGTGGGCTGTGGCTTTGTATCATTTAGGAAGAACCTGTCGACAAGCACAAGAAGC}$ L V G C G F V S F R K K P V D K

#### Figure 11:

a.) Trpl0b cDNA and derived amino acid sequence

```
10
                       30
                                       50
ATGAAATCCTTCCTGTCCACACCATCGTGCTTATCAGGGAGAATGTGTGCAAGTGT
M K S F L P V H T I V L I R E N V C K C
                      90
                                      110
      70
GGCTATGCCCAGAGCCAGCACATGGAAGGCACCCAGATCAACCAAAGTGAGAAATGGAAC
G Y A Q S Q H M E G T Q I N Q S E K W N
                                      170
                      150
     130
TACAAGAACACACCAAGGAATTTCCTACCGACGCCTTTGGGGATATTCAGTTTGAGACA
Y K K H T K E F P T D A F G D I Q F E T
     190
                      210
                                       230
\tt CTGGGGAAGAAGGGAAGTATATACGTCTGTCCTGCGACACGGACGCGGAAATCCTTTAC
L G K K G K Ý I R L S C D T D A E I¨L Y
     250
                      270
                                      290
GAGCTGCTGACCCAGCACTGGCACCTGAAAACACCCCAACCTGGTCATTTCTGTGACCGGG
BLLTQHWHLKTPNLVISVTG
                                       350
     310
                      330
{\tt GGCGCCAAGAACTTCGCCCTGAAGCCGCGCATGCGCAAGATCTTCAGCCGGCTCATCTAC}
G A K N F A L K P R M R K I F S R L I Y
     370
                      390
                                       410
ATCGCGCAGTCCAAAGGTGCTTGGATTCTCACGGGAGGCACCCATTATGGCCTGATGAAG
I A Q S K G A W I L T G G T H Y G L M K
                      450
                                      470
TACATCGGGGAGGTGGTGAGAGATAACACCATCAGCAGGAGTTCAGAGGAGAATATTGTG
Y I G E V V R D N T I S R S S E E N I V
                      510
                                      530
GCCATTGGCATAGCAGCTTGGGGCATGGTCTCCAACCGGGACACCCTCATCAGGAATTGC
A I G I A A W G M V S N R D T L I R N C
                      570
                                      590
GATGCTGAGGGCTATTTTTTAGCCCAGTACCTTATGGATGACTTCACAAGAGATCCACTG
D A E G Y F L A Q Y L M D D F T R D P L
                                      650
                      630
     610
TATATCCTGGACAACCACACACACTTTGCTGCTCGTGGACAATGGCTGTCATGGACAT
Y I L D N N H T H L L L V D N G C H G H
                                      710
     670
                      690
CCCACTGTCGAAGCAAAGCTCCGGAATCAGCTAGAGAAGTATATCTCTGAGCGCACTATT
PTVEAKLRNQLEKYISERTI
                                      770
     730
                      750
CAAGATTCCAACTATGGTGGCAAGATCCCCATTGTGTGTTTTTGCCCAAGGAGGTGGAAAA
Q D S N Y G G K I P I V C F A Q G G G K
     790
                      810
                                      830
GAGACTTTGAAAGCCATCAATACCTCCATCAAAAATAAAATTCCTTGTGTGGTGGTGGAA
E T L K A I N T S I K N K I P C V V V E
                      B70
                                       890
     850
GGCTCGGGCCAGATCGCTGATGTGATCGCTAGCCTGGTGGAGGTGGAGGATGCCCTGACA
G S G Q I A D V I A S L V E V E D A L T
                      930
                                      950
     910
TCTTCTGCCGTCAAGGAGAAGCTGGTGCGCTTTTTACCCCGCACGGTGTCCCGGCTGCCT
S S A V K E K L V R F L P R T V S R L P
                                     1010
     970
                      990
GAGGAGGAGACTGAGAGTTGGATCAAATGGCTCAAAGAAATTCTCGAATGTTCTCACCTA
E E E T E S W I K W L K E I L E C S H L
                     1050
                                      1070
     1030
TTAACAGTTATTAAAATGGAAGAAGCTGGGGATGAAATTGTGAGCAATGCCATCTCCTAC
LTVIKNEEAGDEIVSNAISY
                     1110
                                     1130
     1090
GCTCTATACAAAGCCTTCAGCACCAGTGAGCAAGACAAGGATAACTGGAATGGGCAGCTG
A L Y K A F S T S E Q D K D N W N G Q L
```

Fig. 11 (Continuation)

|      |        | 241  | .0   |                  |       |          |        |            | 2430         | )         |        |              |                 |                                       | 2    | 450      |          |           |           |  |
|------|--------|------|------|------------------|-------|----------|--------|------------|--------------|-----------|--------|--------------|-----------------|---------------------------------------|------|----------|----------|-----------|-----------|--|
| AG   | AAA    | CTTA | .GGZ | \CC              | CAA   | GAT      | TAT    | <b>LAA</b> | GCT          | 3CA       | GAG    | GAT          | GCT             | GAT                                   | CGA' | TGT      | GTT      | CTT       | CTTC      |  |
| R    | N      | L    | G    | Р                | K     | I        | I      | M          | L            | Q         | R      | M            | L               | I                                     | D    | V        | F        | F         | F         |  |
|      |        | 247  | 0    |                  |       |          | 2490   |            |              |           |        |              | 2510            |                                       |      |          |          |           |           |  |
| СТ   | GTT    | CCTC | TTT  | rgc              | GGT   | GTG      | GAT    | GGI        | GGC          | CTT       | TGG    | CGT          | GGC             | CAG                                   | GCA. | AGG      | GAT'     | CCT       | TAGG      |  |
| L    | F      | L    | F    | A                | v     | W        | M      | v          | A            | F         | G      | v            | A               | R                                     | Q    | G        | Ι        | ь         | R         |  |
|      |        | 253  | 0    |                  |       |          |        |            | 2550         | 3         |        |              |                 |                                       | 2    | 570      |          |           |           |  |
| CA   | GAA    | TGAG | CAC  | GCG              | CTG   | GAG      | GTG    | GAT        | ATT          | CCG       | TTC    | GGT          | CAT             | CTA                                   | CGA  | GCC      | CTA      | CCT       | GGCC      |  |
| 0    |        |      |      |                  |       |          |        |            | F            |           |        |              |                 |                                       |      |          |          | L         |           |  |
| ~    |        | 259  | -    |                  |       |          |        |            | 2610         | D         |        |              |                 |                                       | 2    | 630      |          |           |           |  |
| ΔТ   | ርጥጥ    |      | -    | <del>ኒ</del> ርተጥ | GCC   | CAG      | TGA    | CGT        | 'GGA'        | -<br>rgg  | TAC    | CAC          | GTA             | TGA                                   | CTT  | TGC      | CCA      | CTG       | CAC       |  |
| M    |        |      |      |                  | P     |          |        |            | D            |           |        |              |                 |                                       |      |          |          |           |           |  |
| •    | -      | 265  | ~    | •                | -     | _        | _      | -          | 2670         | _         | -      | _            | _               | _                                     |      | 690      |          |           |           |  |
| Trap | יראַרי |      |      | A:YI             | GTC   | מא       | GCC    | ACT        | GTG.         | -<br>ፕረ፣ጥ | CCA    | ദേഷ          | GGA             | TGA                                   | GCA  | CAA      | CCT      | GCC       | CCGC      |  |
|      |        |      |      |                  |       |          |        |            | C            |           |        |              |                 |                                       |      |          |          |           |           |  |
| •    | -      | 271  | -    |                  | ~     | •        | •      | _          | 2736         |           | _      | _            | _               | _                                     |      | 750      | _        | _         |           |  |
| TVT  | 000    |      |      | 2 አጥ             | C7. C | ריא תי   | מממ    | ~~         | GGT          | ~         | ריאים  | מידים        | <sub>ር</sub> ልጥ | بلمانت                                | _    |          | ממי      | СЪТ       | רכזינ     |  |
|      | P      |      |      |                  | T     | - •      |        |            | V            |           |        |              |                 |                                       |      |          |          |           | L         |  |
| r    | P      | 277  |      | 1                | 1     | _        | -      | 11         | 279          | -         | _      | -            | 1.1             | . 11                                  |      | B10      | 14       | -         | _         |  |
| ~    | COM    |      | _    | ~~~              | vo en |          | ር አጥ   | v-m-n      | TGG          | -         | ~~~    | C CT         | cac             | CAC                                   | . –  |          | രവ       | GD D      | רמ מיז    |  |
|      | V      |      |      |                  |       | LGC<br>A |        |            | G            |           |        |              | G               |                                       |      | 0        |          |           | N         |  |
| L    | ٧      | 283  | _    | 1.               | ٧     | A        | 141    | E          | 2850         | _         | 1      | ٧            | G               | 1                                     | •    | 870      | _        | 74        | 14        |  |
| ٠.   |        |      |      | ~~~              |       |          | ~~~    | cama.      | CTT          | -         | VY CIT | ~~~          | aa2             | amy.                                  |      |          |          | a<br>Can  | רח אי     |  |
|      |        |      |      |                  |       |          |        |            | F            |           |        |              |                 |                                       |      |          |          |           |           |  |
| ט    | Q      |      |      | K                | r     | Q        | ĸ      | 1          | 291          |           | v      | Q            | E.              | 1                                     |      | ى<br>930 |          | ш         | 14        |  |
|      | .~~~   | 289  | -    | ~~~              |       |          | amm    | ~~         | TTA          | -         | OT A   | <b>73</b> II | vace            | aam                                   | _    |          |          | (1997)    | רא אר     |  |
|      |        |      |      |                  |       |          |        |            | Y            |           |        |              |                 |                                       |      |          |          |           |           |  |
| Τ.   | P      | _    | -    | F.               | 7     | V        | F      | A          | 2970         |           | x      | IAI          | ٧               | ٧                                     |      | л<br>990 |          | £         | K         |  |
|      |        | 295  |      | ~~~              | a     |          | C17 FF |            | Z970<br>AGTC |           | · ·    | ama          | ama             | · · · · · · · · · · · · · · · · · · · | _    |          |          | אכיא      | ראאי      |  |
|      |        |      |      |                  |       |          |        |            | iGTC:        |           |        |              |                 |                                       |      |          |          |           |           |  |
| C    | С      | -    | K    | E                | K     | N        | M      | E          |              |           | v      | C            | C               | r                                     |      | м<br>050 |          | ע         | 14        |  |
| ~~   | ~~ ~   | 301  |      |                  |       |          | man    | ~~~        | 303)<br>GAA  | -         |        |              | aam             | an cam                                | _    |          |          | CNC       | 7 T T T   |  |
|      |        |      |      |                  |       |          |        |            |              |           |        |              |                 |                                       |      |          |          |           |           |  |
| E    | T      | L    |      | W                | E     | G        | V      | M          | K            |           | N      | ¥            | Į,              | ٧                                     |      |          |          | 1         | I.        |  |
|      |        | 307  | _    | ~~~              |       |          |        | ~ = -      | 309          | -         |        |              | 13 A3           | n cer                                 | _    | 110      |          | ~~~       | י א אודיי |  |
|      |        |      |      | -                |       |          |        |            | GCA'         |           |        |              |                 |                                       |      |          | AAA<br>K | .GC1<br>L | TAA:      |  |
| A    | N      | D    | _    | S                | Е     | B        | M      | R.         | H            |           | F      | R            | Q               | ь                                     | D    | Т        | K        | L         | N         |  |
|      |        | 313  |      |                  |       |          |        |            | 315          | -         |        |              |                 |                                       | _    |          |          |           |           |  |
|      |        |      |      |                  |       |          |        |            | rTGC'        |           |        |              |                 |                                       | Ŀ    |          |          |           |           |  |
| D    | ь      | K    | G    | L                | ь     | ĸ        | E      | I          | Α            | N         | ĸ      | Ι            | к               | *                                     |      |          |          |           |           |  |

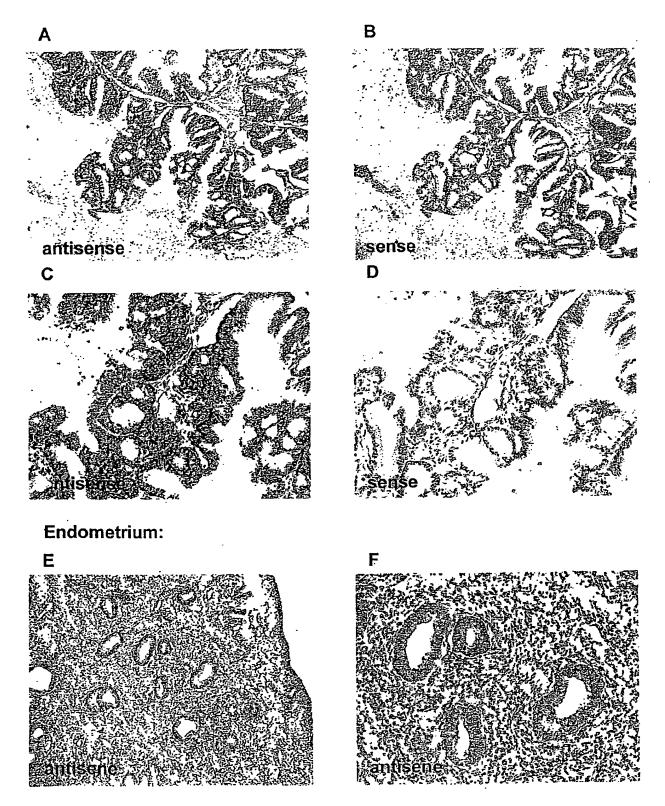
#### b.) Trp10 protein:

MKSFLPVHTIVLIRENVCKCGYAQSQHMEGTQINQSEKWNYKKHTKEFPTDAFGDIQFETLGKKGKYIRLSCDTDAEILY
ELLTQHWHLKTPNLVISVTGGAKNFALKPRMRKIFSRLIYIAQSKGAWILTGGTHYGLMKYIGEVVRDNTISRSSEENIV
AIGIAAWGMVSNRDTLIRNCDAEGYFLAQYLMDDFTRDPLYILDNNHTHLLLVDNGCHGHPTVEAKLRNQLEKYISERTI
QDSNYGGKIPIVCFAQGGGKETLKAINTSIKNKIPCVVVEGSGQIADVIASLVEVEDALTSSAVKEKLVRFLPRTVSRLP
EEETESWIKWLKEILECSHLLTVIKMERAGDEIVSNAISYALYKAFSTSEQDKDNWNGQLKLLLEWNQLDLANDEIFTND
RRWESADLQEVWFTALIKDRPKFVRLFLENGLNIRKFLTHDVLTELFSNHFSTLVYRNLQIAKNSYNDALLTFVWKLVAN
FRRGFRKEDRNGRDEMDIELHDVSPITRHPLQALFIWAILQNKKELSKVIWEQTRGCTLAALGASKLLKTLAKVKNDINA
AGESEELANEYETRAVELFTECYSSDEDLAEQLLVYSCEAWGGSNCLELAVEATDQHFIAQPGVQNPLSKQWYGEISRDT
KNWKIILCLFIIPLVGCGFVSFRKKPVDKHKKLLWYYVAFFTSPFVVFSWNVVFYIAFLLLFAYVLLMDFHSVPHPPELV
LYSLVFVLFCDEVRQWYNNGVNYFTDLWNVMDTLGLFYFIAGIVFRLHSSNKSSLYSGRVIFCLDYIIFTLRLIHIFTVS
RNLGPKIINLQRMLIDVFFFLFLFAVWMVAFGVARQGILRQNEQRWRWIFRSVIYEPYLAMFGQVPSDVDGTTYDFAHCT
FTGNESKPLCVELDEENLPRFPEWITIPLVCIYMLSTNILLVNLLVAMFGYTVGTVQENNDQVWKFQRYFLVQEYCSRLN
IPPFFIVFAYFYMVVKKCFKCCCKEKNMESSVCCFKNEDNETLAWEGVMKENYLVKINTKANDTSEEMRHRFRQLDTKLN
DLKGLLKEIANKIK

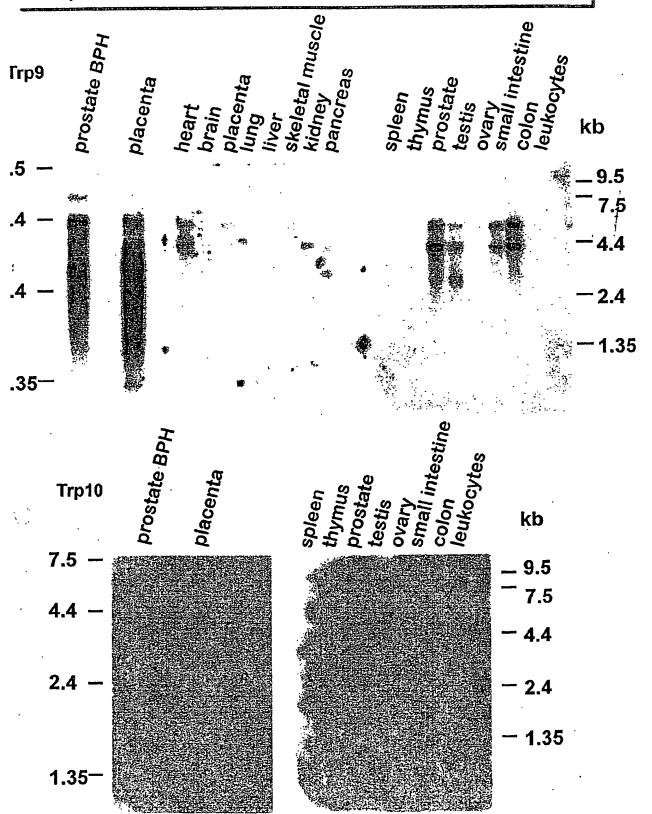
# The Trp8 Gene is expre in normal endometrium

### d in encometrial or uteri

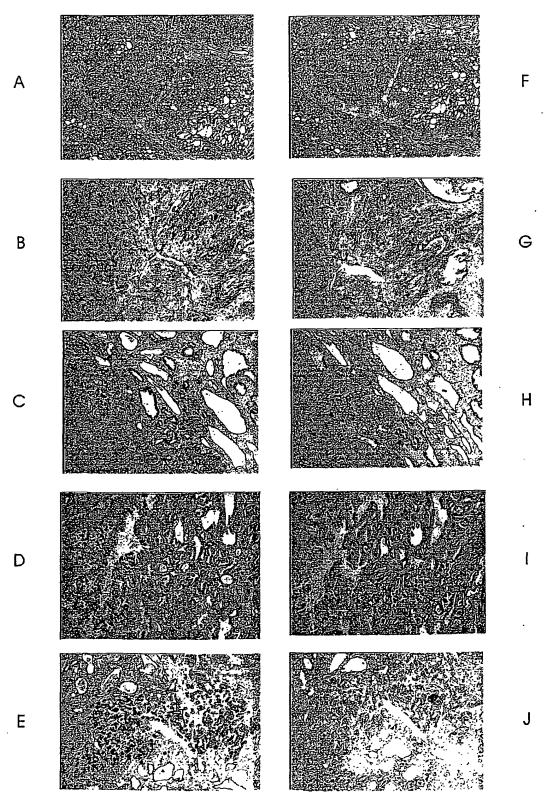
### **Endometrial cancer:**

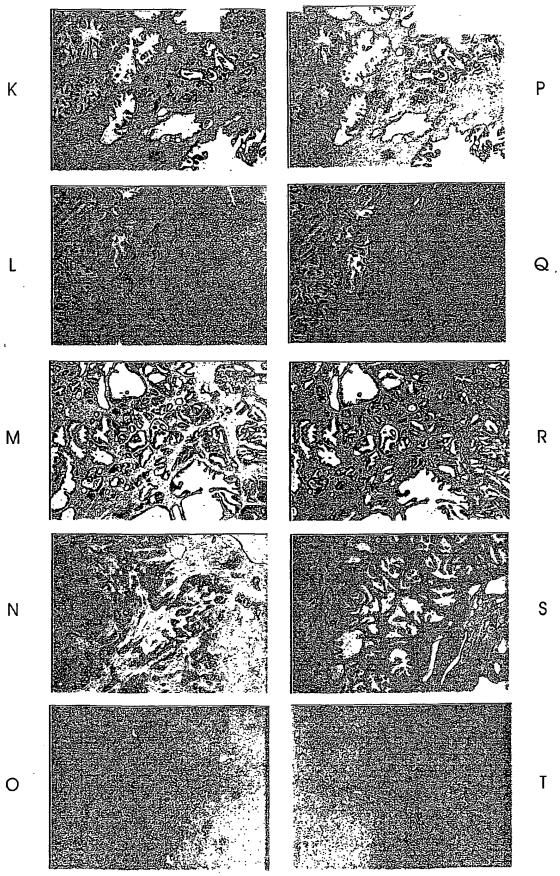


## Expression of human Trp 9 and 1rp 10 |



Expression of Trp10 transcripts and Trp10-antisense transcripts in human prostate cancer and in malignant melanoma





### (19) World Intellectual Property Organization International Bureau





### (43) International Publication Date 7 February 2002 (07.02.2002)

### PCT

### (10) International Publication Number WO 02/010382 A3

- (51) International Patent Classification<sup>7</sup>: C12N 15/12, 15/11, 9/00, C07K 14/47, C12Q 1/68, G01N 33/577, A61K 31/713, C07K 14/705
- (21) International Application Number: PCT/EP01/08309
- (22) International Filing Date: 18 July 2001 (18.07.2001)
- (25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data: 60/221,513

28 July 2000 (28.07.2000) US

- (71) Applicant and
- (72) Inventor: WISSENBACH, Ulrich [DE/DE]; Institut für Pharmakologie und Toxikologie der Uni, versität des Saarlandes, 66421 Homburg (DE).
- (74) Agent: HUBER, Bernard; Huber & Schüssler, Truderinger Str. 246, 81825 München (DE).

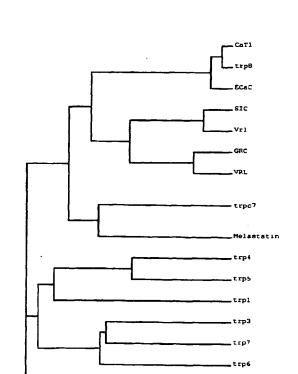
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

#### Published:

with international search report

[Continued on next page]

(54) Title: TRP8, TRP9 AND TRP10, MARKERS FOR CANCER



(57) Abstract: The present invention relates to gene expression in normal cells and cells of malignant tumors and particularly to novel markers associated with cancer, Trp8, Trp9 and Trp10, and the genes encoding Trp8, Trp9 and Trp10. Also provided are vectors, host cells, antibodies, and recombinant methods for producing these human proteins. The invention further relates to diagnostic and therapeutic methods useful for diagnosing and treating a tumor.

WO 02/010382 A3

### WO 02/010382 A3



(88) Date of publication of the international search report: 9 October 2003

t: For two-letter codes and other abbreviations, refer to the "Guid-03 ance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(15) Information about Correction:
Previous Correction:
see PCT Gazette No. 38/2002 of 19 September 2002, Section II

rnational Application No PCT/EP 01/08309

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/12 C12N15/11 C07K14/47 C12Q1/68 C12N9/00 G01N33/577 C07K14/705 A61K31/713 According to International Patent Classification (IPC) or to both national classification and IPC B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC 7 C07K Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the International search (name of data base and, where practical, search terms used) BIOSIS, EPO-Internal, SEQUENCE SEARCH, WPI Data, PAJ C. DOCUMENTS CONSIDERED TO BE RELEVANT Relevant to dalm No. Citation of document, with indication, where appropriate, of the relevant passages Category ° 1-10, WO 99 09166 A (SHAPERO MICHAEL H ; DENDREON Χ 12-17, CORP (US); LAUS REINER (US); TSAVALER) 23,29-31 25 February 1999 (1999-02-25) see \$EQID14 + 15, pages 2,3, 28,29, Example 4 table 3 1-10,12, WO 00 40614 A (BETH ISRAEL HOSPITAL Х ; SCHARENBERG ANDREW M (US)) 13 July 2000 (2000-07-13) see seqid31 + 32, page 11, first paragraph, page 44, lines 13-15 Patent family members are listed in annex. Further documents are listed in the continuation of box C. Special categories of cited documents : "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the "A" document defining the general state of the art which is not considered to be of particular relevance invention "E" earlier document but published on or after the International "X" document of particular relevance; the claimed Invention cannot be considered novel or cannot be considered to Involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed Invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other spectal reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means document published prior to the International filing date but later than the priority date claimed "&" document member of the same patent family Date of mailing of the International search report Date of the actual completion of the international search 13. 03. 2003 6 March 2003 Authorized officer Name and mailing address of the ISA European Patent Office, P.B. 5618 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016

Holtorf, S

national Application No
PCT/EP 01/08309

| C.(Continu | ation) DOCUMENTS CONSIDERED TO BE RELEVANT   |                       |
|------------|--|-----------------------|
| Category ° | Citation of document, with indication, where appropriate, of the relevant passages   | Relevant to claim No. |
| х          | MULLER D ET AL: "Molecular cloning, tissue distribution, and chromosomal mapping of the human epithelial Ca2+ channel (ECAC1)." GENOMICS, vol. 67, no. 1, 1 July 2000 (2000-07-01), pages 48-53, XP002222953 ISSN: 0888-7543 the whole document  | 1                     |
| X          | WO 98 15657 A (ABBOTT LAB)<br>16 April 1998 (1998-04-16)<br>the whole document   | 1-12,<br>29-31        |
| X          | WO 98 37093 A (CORIXA CORP)<br>27 August 1998 (1998-08-27)<br>the whole document   | 1-12,<br>29-31        |
| A          | TSAVALER LARISA ET AL: "TRP-P8, a novel prostate-specific gene, is upregulated in prostate cancer and other malignancies and shares high homology with TRP calcium channel proteins."  PROCEEDINGS OF THE AMERICAN ASSOCIATION FOR CANCER RESEARCH ANNUAL, no. 41, March 2000 (2000-03), page 694 XP008011242  91st Annual Meeting of the American Association for Cancer Research.; San Francisco, California, USA; April 01-05, 2000, March, 2000 ISSN: 0197-016X the whole document |                       |
| <b>A</b>   | HARTENECK C ET AL: "FROM WORM TO MAN:<br>THREE SUBFAMILIES OF TRP CHANNELS"<br>TRENDS IN NEUROSCIENCE, ELSEVIER,<br>AMSTERDAM, NL,<br>vol. 23, no. 4, April 2000 (2000-04),<br>pages 159-166, XP001012870<br>ISSN: 0166-2236   |                       |
| P,X        | WO 01 14423 A (SMITHKLINE BEECHAM PLC)  1 March 2001 (2001-03-01)  see SEQid1 + 2; see example 1  -/   | 1-9,31                |

rnational Application No \*PCT/EP 01/08309

|            | ation) DOCUMENTS CONSIDERED TO BE RELEVANT   | Relevant to daim No.                 |
|------------|--|--------------------------------------|
| Category ° | Citation of document, with Indication, where appropriate, of the relevant passages   | nerevani io gaim No.                 |
| P,X        | WISSENBACH ULRICH ET AL: "Expression of CaT-like, a novel calcium-selective channel, correlates with the malignancy of prostate cancer."  JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 276, no. 22, 1 June 2001 (2001-06-01), pages 19461-19468, XP002222954 ISSN: 0021-9258 the whole document | 1-9,13,<br>14,<br>16-19,<br>21-23,29 |
| P,X        | WO 01 04303 A (HEDIGER MATTHIAS A) 18 January 2001 (2001-01-18) see SEQID1 + 2 the whole document  | 1-5                                  |
| P,X        | WO 01 42467 A (MILLENNIUM PREDICTIVE MEDICINE) 14 June 2001 (2001-06-14) see SEQID 4615  | 1                                    |
| E          | WO 01 51633 A (FANGER GARY RICHARD; HARLOCKER SUSAN L (US); MEAGHER MADELEINE JOY) 19 July 2001 (2001-07-19) see SEQID764, example 3, claims   | 1                                    |
| E          | WO 02 14361 A (AGENSYS INC)<br>21 February 2002 (2002-02-21)<br>see SEQID1479, examples 1-4<br>the whole document  | 1-10,<br>13-23                       |
| E          | WO 02 00722 A (SILOS SANTIAGO INMACULADA<br>;CURTIS RORY A J (US); MILLENNIUM PHARM)<br>3 January 2002 (2002-01-03)<br>see SEQID4  | 1-5                                  |
| E          | WO 01 68857 A (CURTIS RORY A J ;COOK WILLIAM JAMES (US); MILLENNIUM PHARM INC (US) 20 September 2001 (2001-09-20) see SEQID1, examples   | 1-5                                  |
| E          | WO 01 53348 A (SQUIBB BRISTOL MYERS CO; GAUGHAN GLEN T (US); RAMANATHAN CHANDRA S) 26 July 2001 (2001-07-26) see SEQID5 the whole document   | 1                                    |
| E          | WO 01 62794 A (LORA JOSE M ;CURTIS RORY A J (US); GLUCKSMANN MARIA ALEXANDRA (US)) 30 August 2001 (2001-08-30) the whole document  | 1-9                                  |
| E          | WO 02 30268 A (EOS BIOTECHNOLOGY INC) 18 April 2002 (2002-04-18) see SEQID53   | 1,6                                  |
|            | -/   |                                      |

rational Application No PCT/EP 01/08309

|            | ation) DOCUMENTS CONSIDERED TO BE RELEVANT   | Relevant to claim No. |
|------------|--|-----------------------|
| Category ° | Citation of document, with indication, where appropriate, of the relevant passages   | nerevant to ciam No.  |
| T          | BOEDDING MATTHIAS ET AL: "The recombinant human TRPV6 channel functions as Ca2+ sensor in human embryonic kidney and rat basophilic leukemia cells." JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 277, no. 39, 27 September 2002 (2002-09-27), pages 36656-36664, XP002222955 September 27, 2002 ISSN: 0021-9258 the whole document |                       |
|            | ·  |                       |
|            |  |                       |
|            |  |                       |
|            |  |                       |

nternational application No. PCT/EP 01/08309

# INTERNATIONAL SEARCH REPORT

| Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)  |
|--|
| This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:   |
| 1. X Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:   |
| Although claims 24-28 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.  |
| 2. X Claims Nos.: 12 partially because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically: |
| see FURTHER INFORMATION sheet PCT/ISA/210  |
| 3. Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).  |
| Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)  |
| This International Searching Authority found multiple inventions in this international application, as follows:  |
| see additional sheet   |
|  |
| As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.   |
| 2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.  |
| As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:   |
|  |
| 4. No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the Invention first mentioned in the claims; it is covered by claims Nos.:                          |
|  |
| Remark on Protest  The additional search fees were accompanied by the applicant's protest.   |
| No protest accompanied the payment of additional search fees.  |

## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-12, 29-31 partially, 13-28 completely

Isolated nucleic acid molecules encoding human prostate carcinom associated proteins as characterized by SEQIDs5,45,11,3 and SEQIDs 6,46,12,4, respectively; the recombinant expression of the same in host cells; the isolated proteins as characterized by SEQIDs 6,46,12,4; antisense RNA sequence and ribozyme complementary to said nucleic acid molecules; inhibitor that can suppress the activity of said prostate carcinom associated proteins; method for diagnosing a prostate carcinoma by contacting a sample with a nucleic acid, an antibody or other reagent that reacts with the mRNA of SEQIDs5,45,11,3; method for diagnosing endomertial cancer by contacting a target sample with a nucleic acid, an antibody or other reagent that reacts with the mRNA of SEQIDs5,45,11,3; method for diagnosing a melanoma, chorion carcinoma, cancer of the lung and ot the prostate comprising contacting a target sample with a reagent which detects antisense RNA of SEQIDs 11 and 3; method for preventing prostate tumour, endometrial cancer, choroin carcinoma or cancer of the lung comprising administering an inhibiting reagent of human prostate carcinom associated proteins; diagnostic kit containing an antibody; method for identifyng an agonist or an antagonist of human prostate carcinom associated proteins.

### 2. Claims: 1-12, 29-31 partially

Isolated nucleic acid molecule encoding human prostate carcinom associated protein as characterized by SEQIDs 7 and SEQIDs 8, respectively; the recombinant expression of the same in host cells; the isolated protein as characterized by SEQIDs 8; antisense RNA sequence and ribozyme complementary to said nucleic acid molecule; inhibitor that can suppress the activity of said prostate carcinom associated protein; diagnostic kit containing an antibody; method for identifyng an agonist or an antagonist of human prostate carcinom associated proteins.

## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 12 partially

Present claim 12 relates to an inhibitor wich is defined by reference to a desirable characteristic or property, namely suppressing the activity of the protein of claim 6.

The claims cover all inhibitors having this characteristic or property, whereas the application provides only support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT for a limited number of such inhibitors.

In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible. Independent of the above reasoning, the claims also lack clarity (Article 6 PCT). An attempt is made to define the inhibitors by reference to a result to be achieved.

Again, this lack of clarity in the present case is such as to render a meaningful search over the whole of the claimed scope impossible.

Consequently, the search has been carried out for those parts of the claim 12 which appear to be clear, supported and disclosed, namely those parts relating to the Trp8/10 corresponding antibody, Trp8/10 corresponding antisense construct, a Trp8/10 corresponding ribozyme.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

Information on patent family members

rnational Application No
PCT/EP 01/08309

| Patent documer<br>cited in search rep |   | Publication<br>date |  | Patent family<br>member(s)   |  | Publication<br>date  |
|---------------------------------------|---|---------------------|--|--|--|--|
| WO 9909166                            | A | 25-02-1999          | US<br>AU<br>CA<br>EP<br>JP<br>NZ<br>WO               | 6194152   744875   9021898   2300364   1005549   2001514889   503404   9909166   200 | B2<br>A<br>A1<br>A2<br>T<br>A  | 27-02-2001<br>07-03-2002<br>08-03-1999<br>25-02-1999<br>07-06-2000<br>18-09-2001<br>01-03-2002<br>25-02-1999   |
| WO 0040614                            | A | 13-07-2000          | AU<br>CA<br>EP<br>JP<br>WO                           | 2055600 /<br>2360396 /<br>1141017 /<br>2002536966 /<br>0040614 /   | A1<br>A2<br>T  | 24-07-2000<br>13-07-2000<br>10-10-2001<br>05-11-2002<br>13-07-2000   |
| WO 9815657                            | А | 16-04-1998          | US<br>EP<br>JP<br>WO<br>US                           | 5919638 /<br>0954599 /<br>2001523948 /<br>9815657 /<br>6110675 /   | A1<br>T<br>A1  | 06-07-1999<br>10-11-1999<br>27-11-2001<br>16-04-1998<br>29-08-2000   |
| WO 9837093                            | A | 27-08-1998          | USU AUU BRN CZPHOOZPHOOZPHOOZPHOOZPHOOZPHOOZPHOOZPHO | 6261562<br>731840<br>6181898<br>9808881<br>1252837<br>9903016<br>1005546<br>0002095<br>994069<br>337446<br>335348<br>9902053<br>6262245<br>9837093<br>2002090372<br>6465611<br>6395278<br>6329505<br>2002022248<br>2002051977<br>2002193296<br>9801585   | B2<br>A<br>A<br>A<br>T<br>A<br>A<br>A<br>A<br>A<br>A<br>B<br>B<br>B<br>B<br>B<br>B<br>B<br>B<br>B<br>B | 17-07-2001<br>05-04-2001<br>09-09-1998<br>11-09-2001<br>10-05-2000<br>13-03-2002<br>07-06-2000<br>28-10-2000<br>22-10-1999<br>23-02-2001<br>25-04-2000<br>21-04-2000<br>17-07-2001<br>27-08-1998<br>11-07-2002<br>28-05-2002<br>11-12-2001<br>21-02-2002<br>02-05-2002<br>19-12-2002<br>04-09-1998 |
| WO 0114423                            | A | 01-03-2001          | WO   | 0114423  | A1   | 01-03-2001   |
| WO 0104303                            | Α | 18-01-2001          | AU<br>EP<br>WO                                       | 5778600 /<br>1194546 /<br>0104303 /  | A1   | 30-01-2001<br>10-04-2002<br>18-01-2001   |
| WO 0142467                            | Α | 14-06-2001          | AU<br>WO   | 2074201<br>0142467   |  | 18-06-2001<br>14-06-2001   |
| WO 0151633                            | A | 19-07-2001          | AU<br>AU<br>WO<br>EP<br>EP                           | 3447401<br>6158700<br>0104143<br>1194571<br>1261708  | A<br>A2<br>A1  | 24-07-2001<br>30-01-2001<br>18-01-2001<br>10-04-2002<br>04-12-2002   |

Information on patent family members

mational Application No PCT/EP 01/08309

| Patent document<br>cited in search report |   | Publication<br>date |                      | Patent family<br>member(s)  | Publication<br>date  |
|---|---|---------------------|----------------------|---|--|
| WO 0151633                                | A |                     | NO<br>WO<br>US<br>US | 20023402 A<br>0151633 A2<br>2002022248 A1<br>2002051977 A1<br>2002193296 A1 | 29-08-2002<br>. 19-07-2001<br>21-02-2002<br>02-05-2002<br>19-12-2002 |
| WO 0214361                                | Α | 21-02-2002          | AU<br>WO             | 8501801 A<br>0214361 A2   | 25-02-2002<br>21-02-2002   |
| WO 0200722                                | Α | 03-01-2002          | AU<br>WO<br>US       | 7024001 A<br>0200722 A2<br>2002156253 A1                                    | 08-01-2002<br>03-01-2002<br>24-10-2002                               |
| WO 0168857                                | A | 20-09-2001          | AU<br>WO             | 4746001 A<br>0168857 A2   | 24-09-2001<br>20-09-2001   |
| WO 01 <u>5</u> 3348                       | A | 26-07-2001          | AU<br>EP<br>WO<br>US | 3648201 A<br>1252189 A2<br>0153348 A2<br>2002072101 A1                      | 31-07-2001<br>30-10-2002<br>26-07-2001<br>13-06-2002                 |
| WO 0162794                                | Α | 30-08-2001          | AU<br>WO<br>US       | 3859601 A<br>0162794 A2<br>2002142377 A1                                    | 03-09-2001<br>30-08-2001<br>03-10-2002                               |
| WO 0230268                                | Α | 18-04-2002          | US<br>AU<br>WO       | 2002068036 A1<br>1534502 A<br>0230268 A2                                    | 06-06-2002<br>22-04-2002<br>18-04-2002                               |

# CORRECTED VERSION

## (19) World Intellectual Property Organization International Bureau





(43) International Publication Date 7 February 2002 (07.02.2002)

**PCT** 

## (10) International Publication Number WO 02/010382 A2

- (51) International Patent Classification7: C12N 15/12. 15/11, 9/00, C07K 14/47, C12Q 1/68, G01N 33/577, A61K 31/713
- (21) International Application Number: PCT/EP01/08309
- (22) International Filing Date: 18 July 2001 (18.07.2001)
- (25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data: 60/221,513

28 July 2000 (28.07.2000) US

(71) Applicant and

- (72) Inventor: WISSENBACH, Ulrich [DE/DE]; Institut für Pharmakologie und Toxikologie der Uni, versität des Saarlandes, 66421 Homburg (DE).
- (74) Agent: HUBER, Bernard; Huber & Schüssler, Truderinger Str. 246, 81825 München (DE).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,

LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

#### Published:

- without international search report and to be republished upon receipt of that report
- (48) Date of publication of this corrected version:

19 September 2002

(15) Information about Correction:

see PCT Gazette No. 38/2002 of 19 September 2002, Section II

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: TRP8, TRP9 AND TRP10, NOVEL MARKERS FOR CANCER

(57) Abstract: The present invention relates to gene expression in normal cells and cells of malignant tumors and particularly to novel markers associated with cancer, Trp8, Trp9 and Trp10, and the genes encoding Trp8, Trp9 and Trp10. Also provided are vectors, host cells, antibodies, and recombinant methods for producing these human proteins. The invention further relates to diagnostic and therapeutic methods useful for diagnosing and treating a tumor.

## Trp8, Trp9 and Trp10, novel markers for cancer

#### FIELD OF THE INVENTION

The present invention relates to gene expression in normal cells and cells of malignant tumors and particularly to novel markers associated with cancer, Trp8, Trp9 and Trp10, and the genes encoding Trp8, Trp9 and Trp10

## **BACKROUND OF THE TECHNOLOGY**

Prostate cancer is one of the most common diseases of older men world wide. Diagnosis and monitoring of prostate cancer is difficult because of the heterogeneity of the disease. For diagnosis different grades of malignancy can be distinguished according to the Gleason-Score Diagnosis. For this diagnosis a prostate tissue sample is taken from the patient by biopsy and the morphology of the tissue is investigated. However, this approach only yields subjective results depending on the experience of the pathologist. For confirmation of these results and for obtaining an early diagnosis an additional diagnostic method can be applied which is based on the detection of a prostate specific antigen (PSA). PSA is assayed in serum samples, blood samples etc. using an anti-PSA-antibody. However, since in principle PSA is also expressed in normal prostate tissue there is a requirement for the definition of a threshold value (about 4 ng/ml PSA) in order to be able to distinguish between normal and malign prostate tissue. Unfortunately, this diagnostic method is quite insensitive and often yields false-positive results. Moreover, by using this diagnostic method any conclusions as regards the grade of malignancy, the progression of the tumor and its potential for metastasizing cannot be drawn. Thus, the use of molecular markers would be helpful to distinguish benign from malign tissue and for grading and staging prostate carcinoma, particularly for patients with metastasizing prostate cancer having a very bad prognosis.

The above discussed limitations and failings of the prior art to provide meaningful specific markers which correlate with the presence of prostate tumors, in particular metastasizing tumors, has created a need for markers which can be used diagnostically, prognostically and therapeutically over the course of this disease. The present invention fulfils such a need by the provision of Tpr8, Trp9 and Trp10 and the genes encoding Trp8, Trp9 and Trp10: The genes encoding Trp8 and Trp10 are expressed in prostate carcinoma and prostatic metastasis, but

1

not in normal prostate, benign hyperplasia (BHP) and intraepithelial prostatic neoplasia (PIN). Furthermore, expression of Trp10 transcripts is detectable in carcinoma but not in healthy tissue of the lung, the prostate, the placenta and in melanoma.

## SUMMARY OF THE INVENTION

The present invention is based on the isolation of genes encoding novel markers associated with a cancer, Trp8, Trp9 and Trp10. The new calcium channel proteins Trp8, Trp9 and Trp10 are members of the trp (transient receptor potential) - family, isolated from human placenta (Trp8a and Trp8b) and humane prostate (Trp9, Trp10a and Trp10b). Trp proteins belong to a steadily growing family of Ca2+ selective and non selective ion channels. In the recent years seven Trp proteins (trp1 - trp7) have been identified and suggested to be involved in cation entry, receptor operated calcium entry and pheromone sensory signaling. Structurally related to the trp proteins are the vanilloid receptor (VR1) and the vanilloid like receptor (VRL-1) both involved in nociception triggered by heat. Furthermore, two calcium permeable channels were identified in rat small intestine (CaT1) and rabbit kidney (ECaC). These distantly related channels are suggested to be involved in the uptake of calcium ions from the lumen of the small intestine (CaT1) or in the reuptake of calcium ions in the distal tubule of the kidney (ECaC). Common features or the Trp and related channels are a proposed structure comprising six transmembrane domains including several conserved amino acid motifs. In the present invention the cloning and expression of a CaT1 like calcium channel (Trp8) from human placenta as well as Trp9 and Trp10 (two variants, Trp10a and Trp10b) is described. Two polymorphic variants of the Trp8 cDNA were isolated from placenta (Trp8a and Trp8b). Transient expression of the Trp8b cDNA in HEK (human embryonic kidney) cells results in cytosolic calcium overload implicating that the Trp8 channel is constitutive open in the expression system. Trp8 induces highly calcium selective inward currents in HEK cells. The C-terminus of the Trp8 protein binds calmodulin in a calcium dependent manner. The Trp9 channel is expressed in trophoblasts and syncytiotrophoblasts of placenta and in pancreatic acinar cells. Furthermore, the Trp8 channel is expressed in prostatic carcinoma and prostatic metastases, but not in normal tissue of the prostate. No expression of Trp8 transcripts is detectable in benign prostatic hyperplasia (BPH) or prostatic intraepithelial neoplasia (PIN). Therefore, the Trp8 channel is exclusively expressed in malign prostatic tissues and serves as molecular marker for prostate cancer. From the experimental results it is also apparent that the

modulation of Trp8 and/or Trp10, e.g. the inhibition of expression or activity, is of therapeutic interest, e.g. for the prevention of tumor progression.

The present invention, thus, provides a Trp8, Trp9 and Trp10 protein, respectively, as well as nucleic acid molecule encoding the protein and, moreover, an antisense RNA, a ribozyme and an inhibitor, which allow to inhibit the expression or the activity of Trp8, Trp9 and/or Trp10.

In one embodiment, the present invention provides a diagnostic method for detecting a prostate cancer or endometrial cancer (cancer of the uterus) associated with Trp8 or Trp10 in a tissue of a subject, comprising contacting a sample containing Trp8 and/or Trp10 encoding mRNA with a reagent which detects Trp8 and/or Trp10 or the corresponding mRNA.

In a further embodiment, the present invention provides a diagnostic method for detecting a melanoma, chorion carcinoma, cancer of the lung and of the prostate in a tissue of a subject, comprising contacting a sample with a reagent which detects Trp10a and/or Trp10b antisense transcripts or Trp10a and/or Trp10b related antisense transcripts.

In another embodiment, the present invention provides a method of treating a prostate tumor, carcinoma of the lung, carcinoma of the placenta (chorion carcinoma) or melanoma associated with Trp8 and/or Trp10, comprising administering to a subject with such an disorder a therapeutically effect amount of a reagent which modulates, e.g. inhibits, expression of Trp8 and/or Trp10 or the activity of the protein, e.g. the above described compounds.

Finally, the present invention provides a method of gene therapy comprising introducing into cells of a subject an expression vector comprising a nucleotide sequence encoding the above mentioned antisense RNA or ribozyme, in operable linkage with a promoter.

#### BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1: A, phylogenetic relationship of trp and related proteins. B, hydropathy plot of the Trp8 protein sequence according to Kyte and Doolittle. C, alignment of Trp8a/b to the epithelial calcium channels ECaC (from rabbit) and Vr1 (from rat). Putative transmembrane domains are underlined.

Figure 2: A, polymorphism of the Trp8 gene. The polymorphic variants Trp8a and Trp8b differ in five base pairs resulting in three amino acid exchanges in the derived protein sequences. Specific primers were derived from the Trp8 gene as indicated by arrows. B, the Trp8a and Trp8b genes are distinguishable by a single restriction site. Genomic fragments of the Trp8 gene can be amplified using specific primers (shown in A). The genomic fragment of the Trp8b gene contains an additional site of the restriction enzyme BSP1286I (B). C, the Trp8 gene is located on chromosome 7. D, genotyping of eleven human subjects. A 458 bp genomic fragment of the Trp8 gene was amplified using specific primers (shown in A) and restricted with BSP1286I. The resulting fragments were analyzed by PAGE electrophoresis.

Figure 3: The Trp8b protein is a calcium selective ion channel. A, representative trace of a pdiTrp8b transfected HEK 293 cell. Trp8b mediated currents are activated by voltage ramps (-100 mV - +100 mV) of 100 msec at -40 mV or +70 mV holding potential. 1, Trp8b currents in the presence at 2mm  $[Ca^{2+}]_0$ ; 2, effect of solution switch alone 3, switch to nominal zero calcium solution. B, Trp8b currents in the presence of zero divalent cations. C, current voltage relationship of the currents shown in A. Inset, leak subtracted current. D, current voltage relationship of the current shown in B. E, statistics of representative experiments. Black: Trp8 transfected cells, gray: control cells. Columns from left to right: Trp8 currents at -40 mV (n=12) and +70 mV holding potential (n=12). Trp8 currents in standard bath solution including 120 mM NMDG without sodium (n=7) and with nominal zero calcium ions (n=8) or in the presence of 1mM EGTA with zero divalent cations (n=6). F, representative changes in  $[Ca^{2+}]_i$  in Trp8b transfected HEK cells (gray) and controls (black) in the presence or absence of 1mM  $[Ca^{2+}]_0$ . Inset, relative increase of cytosolic calcium concentration of Trp8b transfected HEK cells, before and after readdition of 1 mM  $[Ca^{2+}]_0$  in comparison to control cells.

<u>Figure 4</u>: The C-terminal region of the Trp8 protein binds calmodulin. A, N- and C-terminal fragments of the Trp8 protein used for calmodulin binding studies. B, the Trp8 protein and a truncated Trp8 protein which was in vitro translated after MunI cut of the cDNA, which lacks the C-terminal 32 amino acid residues, were in vitro translated in the presence of <sup>35</sup>S-methionine and incubated with calmodulin coupled agarose beads in the presence of 1 mM Ca<sup>2+</sup> or 2 mM EGTA. C, calmodulin binding to N- and C-terminal fragments of the Trp8protein in the presence of Ca<sup>2+</sup> (1 mM) or EGTA (2 mM)

Figure 5: Expression pattern of the Trp8 cDNA. A, Northern blots (left panels, Clontech, Palo Alto) were hybridized using a 348 bp NcoI/BamHT fragment of the Trp9 cDNA. The probe hybridizes to mRNA species isolated from the commercial blot, but not to mRNA species isolated from benign prostate hyperplasia (right panel, mRNA isolated from 20 human subjects with benign prostate hyperplasia). B,C, in situ hybridization with biotinylated Trp8 specific oligonucleotides on slides of human tissues. Left column antisense probes, right column sense probes. D, antinsense probes.

<u>Figure 6:</u> Differential expression of Trp8 cDNA in human prostate. A-F, in situ hybridization with prostatic tissues. A, normal prostate, B, primary carcinoma, C, benign hyperplasia, D, rezidive carcinoma, E, prostatic intraepithelial neoplasia, F, lymphnode metastasis of the prostata.

Figure 7: Trp8a cDNA sequence and derived amino acid sequence

Figure 8: A, Trp8b cDNA sequence and derived amino acid sequence

B, cDNA sequence of splice variant 1 (12B1)

C, cDNA sequence of splice variant 2 (17-3)

D, cDNA sequence of splice variant 3 (23A3)

E, cDNA sequence of splice variant 4 (23C3)

<u>Figure 9:</u> A, Trp9 cDNA sequence and derived amino acid sequence B, cDNA sequence of splice variant 15 and derived amino acid sequence.

<u>Figure 10:</u> A, cDNA sequence of Trp10a and derived amino acid sequence, B, cDNA fragment of Trp10a and derived amino acid sequence.

Figure 11: cDNA sequence of Trp10b and derived amino acid sequence.

Figure 12: Expression of Trp8 mRNA in human endometrial cancer or cancer of the uterus. A - D, in situ hybridization with slides of endometrial cancer hybridized with Trp8 antisense (left column) or sense probes as controls (right column). E - F, Trp8 antisense probes hybridized to slides of normal endometrium. It can be clearly seen no hybridization occurs with normal endometrial tissue.

## Figure 13: Expression of human Trp9 and Trp10 genes

Northern blots were hybridized using Trp9 (upper panel) or Trp10 (lower panel) specific probes. Expression of the Trp9 cDNA is detectable in many tissues including human prostate and colon as well as in benign prostatic hyperplasia. Expression of Trp10 cDNA is detectable in human prostate of a commercial northern blot (Clontech, right side). This Northern blot contains prostatic tissue collected from 15 human subjects in the range of 14 - 60 years of age. No expression of Trp10 cDNA was detectable in benign prostatic hyperplasia (left side).

Figure 14: Expression of Trp10 transcripts and Trp10-antisense transcripts in human prostate cancer and metastasis of a melanoma. In situ hybridizations of slides hybridized with Trp10-antisense (A-E, K-N) and Trp10 related sense probes (F-J, P-R). It can clearly be seen that both probes detect the same cancer cells indicating that these cancer cells express Trp10 transcripts as well as Trp10-antisense transcripts. S, no Trp10 expression is detectable in benign hyperplasia of the prostate (BPH). O and T, show expression of Trp10 transcripts (O) and Trp10-antisense transcripts (T) in a metastasis of a melanoma in human lung. Melanoma cancer cells express both Trp10 transcripts and Trp10-antisense transcripts.

## DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to an isolated nucleic acid molecule encoding the human prostate carcinoma associated protein Trp8a, Trp8b, Trp9, Trp10a or Trp10b or a protein exhibiting biological properties of Trp8a, Trp8b, Trp9, Trp10a or Trp10b and being selected from the group consisting of

- (a) a nucleic acid molecule encoding a protein that comprises the amino acid sequence depicted in Figure 7, 8A, 9,10 or 11;
- (b) a nucleic acid molecule comprising the nucleotide sequence depicted in Figure 7, 8A, 9,10, or 11;
- (c) a nucleic acid molecule included in DSMZ Deposit no. DSM 13579 (deposit date: 28 June 2000), DSM 13580 (deposit date: 28 June 2000), DSM 13584 (deposit date: 5 July 2000), DSM 13581 (deposit date: 28 June 2000) or DSM ....(deposit date:....);
- (d) a nucleic acid molecule with hybridizes to a nucleic acid molecule specified in (a) to(c)

(e) a nucleic acid molecule the nucleic acid sequence of which deviates from the nucleic sequences specified in (a) to (d) due to the degeneration of the genetic code; and

(f) a nucleic acid molecule, which represents a fragment, derivative or allelic variation of a nucleic acid sequence specified in (a) to (e).

As used herein, a protein exhibiting biological properties of Trp8a, Trp8b, Trp9,Trp10a or Trp10b is understood to be a protein having at least one of the activities as illustrated in the Examples, below.

As used herein, the term "isolated nucleic acid molecule, includes nucleic acid molecules substantially free of other nucleic acids, proteins, lipids, carbohydrates or other materials with which it is naturally associated.

In a first embodiment, the invention provides an isolated nucleic acid molecule encoding the human prostate carcinoma associated protein Trp8a, Trp8b, Trp9,Trp10a or Trp10b comprising the amino acid sequence depicted in Figure 7, 8A, 9,10 or 11. The present invention also provides a nucleic acid molecule comprising the nucleotide sequence depicted in Figure 7, 8A, 9,10 or 11.

The present invention provides not only the generated nucleotide sequence identified in Figure 7, 8A, 9,10 or 11, respectively and the predicted translated amino acid sequence, respectively, but also plasmid DNA containing a Trp8a cDNA deposited with the DSMZ, under DSM 13579, a Trp8b cDNA deposited with the DSMZ, under DSM 13580, a Trp9 cDNA deposited with the DSMZ, under DSM 13581, and a Trp10b cDNA deposited with the DSMZ, under DSM...., respectively. The nucleotide sequence of each deposited Trp-clone can readily be determined by sequencing the deposited clone in accordance with known methods. The predicted amino acid sequence can then be verified from such deposits. Moreover, the amino acid sequence of the protein encoded by each deposited clone can also be directly determined by peptide sequencing or by expressing the protein in a suitable host cell containing the deposited Trp-encoding DNA, collecting the protein, and determining its sequence.

The nucleic acid molecules of the invention can be both DNA and RNA molecules. Suitable DNA molecules are, for example, genomic or cDNA molecules. It is understood that all

nucleic acid molecules encoding all or a portion of Trp8a, Trp8b, Trp9,Trp10a or Trp10b are also included, as long as they encode a polypeptide with biological activity. The nucleic acid molecules of the invention an be isolated from natural sources or can be synthesized according to know methods.

The present invention also provides nucleic acid molecules which hybridize to the above nucleic acid molecules. As used herein, the term "hybridize,, has the meaning of hybridization under conventional hybridization conditions, preferably under stringent conditions as described, for example, in Sambrook et al., Molecular Cloning, A Laboratory Manual 2<sup>nd</sup> edition (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. Also contemplated are nucleic acid molecules that hybridize to the Trp nucleic acid molecules at lower stringency hybridization conditions. Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages of formamide result in lowered stringency), salt conditions, or temperature. For example, lower stringency conditions include an overnight incubation at 37°Cin a solution comprising 6X SSPE (20X SSPE = 3M NaCl; 9.2M NaH<sub>2</sub>PO<sub>4</sub>; 0.02M EDTA, pH7.4), 0.5% SDS, 30% formamide, 100 µg/ml salmon sperm blocking DNA, following by washes at 50°C with 1 X SSPE, 0.1% SDS. In addition, to achieve even lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5X SSC). Variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

Nucleic acid molecules that hybridize to the molecules of the invention can be isolated, e.g., from genomic or cDNA libraries that were produced from human cell lines or tissues. In order to identify and isolate such nucleic acid molecules the molecules of the invention or parts of these molecules or the reverse complements of these molecules can be used, for example by means of hybridization according to conventional methods (see, e.g., Sambrook et al., supra). As a hybridization probe nucleic acid molecules can be used, for example, that have exactly or basically the nucleotide sequence depicted in Figure 7, 8A, 9,10 or 11, respectively, or parts of these sequences. The fragments used as hybridization probe can be synthetic

fragments that were produced by means of conventional synthetic methods and the sequence of which basically corresponds to the sequence of a nucleic acid molecule of the invention.

The nucleic acid molecules of the present invention also include molecules with sequences that are degenerate as a result of the genetic code.

In a further embodiment, the present invention provides nucleic acid molecules which comprise fragments, derivatives and allelic variants of the nucleic acid molecules described above encoding a protein of the invention. "Fragments,, are understood to be parts of the nucleic acid molecules that are long enough to encode one of the described proteins. These fragments comprise nucleic acid molecules specifically hybridizing to transcripts of the nucleic acid molecules of the invention. These nucleic acid molecules can be used, for example, as probes or primers in the diagnostic assay and/or kit described below and, preferably, are oligonucleotides having a length of at least 10, in particular of at least 15 and particularly preferred of at least 50 nucleotides. The nucleic acid molecules and oligonucleotides of the invention can also be used, for example, as primers for a PCR reaction. Examples of particular useful probes (primers) are shown in Tables 1 and 2.

#### Table 1

Trp8 probes used for in situ hybridization:

Probes (antisense)

- 1.) 5' TCCGCTGCCGGTTGAGATCTTGCC 3'
- 2.) 5' CTTGCTCCATAGGCAGAGAATTAG 3'
- 3.) 5' ATCCTCAGAGCCCCGGGTGTGGAA3'

#### Controls (sense)

- 1.) 5' GGCAAGATCTCAACCGGCAGCGGA 3'
- 2.) 5' CTAATTCTCTGCCTATGGAGCAAG 3'
- 3.) 5' TTCCACACCCGGGGCTCTGAGGAT 3'

## Tabelle 2

Trp10 probes used for the in situ hybridizations shown in Figure 14:

## Probes (antisense)

1.) 5' GCTTCCACCCCAAGCTTCACAGGAATAGA 3' (Figure 14 A, 14B)

2.) 5' GGCGATGAAATGCTGGTCTGTGGC 3' (Figure 14C, 14D, 14N, 14S, 14O)

3.) 5' ATCTTCCAGTTCTTGGTGTCTCGG 3' (Figure 14E, 14K)

4.) 5' GCTGCAGTACTCCTGCACCAGGAA 3' (Figure 14L, 14M)

## Probes (sense)

1.) 5' TCTATTCCTGTGAAGCTTGGGGTGGAAGC 3' (Figure 14F, 14G)

2.) 5' GCCACAGACCAGCATTTCATCGCC 3' (Figure 14H, 14I, 14T)

3.) 5' CCGAGACACCAAGAACTGGAAGAT 3' (Figure 14J, 14P)

4.) 5' TTCCTGGTGCAGGAGTACTGCAGC 3' (Figure 14Q, 14R)

The term "derivative,, in this context means that the sequences of these molecules differ from the sequences of the nucleic acid molecules described above at one or several positions but have a high level of homology to these sequences. Homology hereby means a sequence identity of at least 40%, in particular an identity of at least 60%, preferably of more than 80% and particularly preferred of more than 90%. These proteins encoded by the nucleic acid molecules have a sequence identity to the amino acid sequence depicted in Figure 7, 8A, 9, 10 and 11, respectively, of at least 80%, preferably of 85% and particularly preferred of more than 90%, 97% and 99%. The deviations to the above-described nucleic acid molecules may have been produced by deletion, substitution, insertion or recombination. The definition of the derivatives also includes splice variants, e.g. the splice variants shown in Figures 8B to 8E and 9B.

The nucleic acid molecules that are homologous to the above-described molecules and that represent derivatives of these molecules usually are variations of these molecules that represent modifications having the same biological function. They can be naturally occurring variations, for example sequences from other organisms, or mutations that can either occur naturally or that have been introduced by specific mutagenesis. Furthermore the variations can be synthetically produced sequences. The allelic variants can be either naturally occurring variants or synthetically produced variants or variants produced by recombinant DNA processes.

Generally, by means of conventional molecular biological processes it is possible (see, e.g., Sambrook et al., supra) to introduce different mutations into the nucleic acid molecules of the invention. As a result Trp proteins or Trp related proteins with possibly modified biological properties are synthesized. One possibility is the production of deletion mutants in which nucleic acid molecules are produced by continuous deletions from the 5'- or 3'-terminal of the coding DNA sequence and that lead to the synthesis of proteins that are shortened accordingly. Another possibility is the introduction of single-point mutation at positions where a modification of the amino aid sequence influences, e.g., the ion channel properties or the regulations of the trp-ion channel. By this method muteins can be produced, for example, that possess a modified ion conducting pore, a modified K<sub>m</sub>-value or that are no longer subject to the regulation mechanisms that normally exist in the cell, e.g. with regard to allosteric regulation or covalent modification. Such muteins might also be valuable as therapeutically useful antagonists of Trp8a, Trp8b, Trp9, Trp10a or Trp10b, respectively.

For the manipulation in prokaryotic cells by means of genetic engineering the nucleic acid molecules of the invention or parts of these molecules can be introduced into plasmids allowing a mutagenesis or a modification of a sequence by recombination of DNA sequences. By means of conventional methods (cf. Sambrook et al., supra) bases can be exchanged and natural or synthetic sequences can be added. In order to link the DNA fragments with each other adapters or linkers can be added to the fragments. Furthermore, manipulations can be performed that provide suitable cleavage sites or that remove superfluous DNA or cleavage sites. If insertions, deletions or substitutions are possible, in vitro mutagenesis, primer repair, restriction or ligation can be performed. As analysis method usually sequence analysis, restriction analysis and other biochemical or molecular biological methods are used.

The proteins encoded by the various variants of the nucleic acid molecules of the invention show certain common characteristics, such as ion channel activity, molecular weight, immunological reactivity or conformation or physical properties like the electrophoretical mobilty, chromatographic behavior, sedimentation coefficients, solubility, spectroscopic properties, stability; pH optimum, temperature optimum.

The invention furthermore relates to vectors containing the nucleic acid molecules of the invention. Preferably, they are plasmids, cosmids, viruses, bacteriophages and other vectors

usually used in the field of genetic engineering. Vectors suitable for use in the present invention include, but are not limited to the T7-based expression vector for expression in mammalian cells and baculovirus-derived vectors for expression in insect cells. Preferably, the nucleic acid molecule of the invention is operatively linked to the regulatory elements in the recombinant vector of the invention that guarantee the transcription and synthesis of an RNA in prokryotic and/or eukaryotic cells that can be translated. The nucleotide sequence to be transcribed can be operably linked to a promoter like a T7, metallothionein I or polyhedrin promoter.

In a further embodiment, the present invention relates to recombinant host cells transiently or stable containing the nucleic acid molecules or vectors or the invention. A host cell is understood to be an organism that is capable to take up *in vitro* recombinant DNA and, if the case may be, to synthesize the proteins encoded by the nucleic acid molecules of the invention. Preferably, these cells are prokaryotic or eukaryotic cells, for example mammalian cells, bacterial cells, insect cells or yeast cells. The host cells of the invention are preferably characterized by the fact that the introduced nucleic acid molecule of the invention either is heterologous with regard to the transformed cell, i.e. that it does not naturally occur in these cells, or is localized at a place in the genome different from that of the corresponding naturally occurring sequence.

A further embodiment of the invention relates to isolated proteins exhibiting biological properties of the human prostate carcinoma associated protein Trp8a, Trp8b, Trp9,Trp10a or Trp10b and being encoded by the nucleic acid molecules of the invention, as well as to methods for their production, whereby, e.g., a host cell of the invention is cultivated under conditions allowing the synthesis of the protein and the protein is subsequently isolated from the cultivated cells and/or the culture medium. Isolation and purification of the recombinantly produced proteins may be carried out by conventional means including preparative chromatography and affinity and immunological separations involving affinity with an anti-Trp8a-, anti-Trp8b-, anti-Trp9-,anti-Trp10a- or anti-Trp10b-antibody, respectively.

As used herein, the term "isolated protein, includes proteins substantially free of other proteins, nucleic acids, lipids, carbohydrates or other materials with which it is naturally associated. Such proteins however not only comprise recombinantly produced proteins but include isolated naturally occurring proteins, synthetically produced proteins, or proteins

produced by a combination of these methods. Means for preparing such proteins are well understood in the art. The Trp proteins are preferably in a substantially purified form. A recombinantly produced version of a human prostate carcinoma associated protein Trp8a, Trp8b, Trp9,Trp10a or Trp10b protein, including the secreted protein, can be substantially purified by the one-step method described in Smith and Johnson, Gene 67; 31-40 (1988).

In a further preferred embodiment, the present invention relates to an antisense RNA sequence characterised that it is complementary to an mRNA transcribed from a nucleic acid molecule of the present invention or a part thereof and can selectively bind to said mRNA, said sequence being capable of inhibiting the synthesis of the protein encoded by said nucleic acid molecules, and a ribozyme characterised in that it is complementary to an mRNA transcribed from a nucleic acid molecule of the present invention or a part thereof and can selectively bind to and cleave said mRNA, thus inhibiting the synthesis of the proteins encoded by said nucleic acid molecules. Riboyzmes which are composed of a single RNA chain are RNA enzymes, i.e. catalytic RNAs, which can intermolecularly cleave a target RNA, for example the mRNA transcribed from one of the Trp genes. It is now possible to construct ribozymes which are able to cleave the target RNA at a specific site by following the strategies described in the literature. (see, e.g., Tanner et al., in: Antisense Research and Applications, CRC Press Inc. (1993), 415-426). The two main requirements for such ribozymes are the catalytic domain and regions which are complementary to the target RNA and which allow them to bind to its substrate, which is a prerequisite for cleavage. Said complementary sequences, i.e., the antisense RNA or ribozyme, are useful for repression of Trp8a-, Trp8b, Trp9-,Trp10a- and Trp10b-expression, respectively, i.e. in the case of the treatment of a prostate cancer or endometrial cancer (carcinoma of the uterus). Preferably, the antisense RNA and ribozyme of the invention are complementary to the coding region. The person skilled in the art provided with the sequences of the nucleic acid molecules of the present invention will be in a position to produce and utilise the above described antisense RNAs or ribozymes. The region of the antisense RNA and ribozyme, respectively, which shows complementarity to the mRNA transcribed from the nucleic acid molecules of the present invention preferably has a length of at least 10, in particular of at least 15 and particularly preferred of at least 50 nucleotides.

In still a further embodiment, the present invention relates to inhibitors of Trp8a, Trp8b, Trp9, Trp10a and Trp10b, respectively, which fulfill a similar purpose as the antisense RNAs or

ribozymes mentioned above, i.e. reduction or elimination of biologically active Trp8a, Trp8b, Trp9, Trp10a or Trp10b molecules. Such inhibitors can be, for instance, structural analogues of the corresponding protein that act as antagonists. In addition, such inhibitors comprise molecules identified by the use of the recombinantly produced proteins, e.g. the recombinantly produces protein can be used to screen for and identify inhibitors, for example, by exploiting the capability of potential inhibitors to bind to the protein under appropriate conditions. The inhibitors can, for example, be identified by preparing a test mixture wherein the inhibitor candidate is incubated with Trp8a, Trp8b, Trp9, Trp10a or Trp10b, respectively, under appropriate conditions that allow Trp8a, Trp8b, Trp9, Trp10a or Trp10b to be in a native conformation. Such an in vitro test system can be established according to methods well known in the art. Inhibitors can be identified, for example, by first screening for either synthetic or naturally occurring molecules that bind to the recombinantly produced Trp protein and then, in a second step, by testing those selected molecules in cellular assays for inhibition of the Trp protein, as reflected by inhibition of at least one of the biological activities as described in the examples, below. Such screening for molecules that bind Trp8a, Trp8b, Trp9, Trp10a or Trp10b could easily performed on a large scale, e.g. by screening candidate molecules from libraries of synthetic and/or natural molecules. Such an inhibitor is, e.g., a synthetic organic chemical, a natural fermentation product, a substance extracted from a microorganism, plant or animal, or a peptide. Additional examples of inhibitors are specific antibodies, preferably monoclonal antibodies. Moreover, the nucleic sequences of the invention and the encoded proteins can be used to identify further factors involved in tumor development and progression. In this context it should be emphasized that the modulation of the calcium channel of a member of the trp family can result in the stimulation of the immune response of T lymphocytes leading to proliferation of the T lymphocytes. The proteins of the invention can, e.g., be used to identify further (unrelated) proteins which are associated with the tumor using screening methods based on protein/protein interactions, e.g. the two-hybridsystem Fields, S. and Song, O. (1989) Nature (340): 245-246.

The present invention also provides a method for diagnosing a prostate carcinoma which comprises contacting a target sample suspected to contain the protein Trp8a, Trp8b, Trp10a and/or Trp10b or the Trp8a, Trp8b, Trp10a and/or Trp10b encoding mRNA with a reagent which reacts with Trp8a, Trp8b, Trp10a and/or Trp10b or the Trp8a, Trp8b, Trp10a and/or Trp10b encoding mRNA and detecting Trp8a, Trp8b, Trp10a and/or Trp10b or Trp8a, Trp8b, Trp10a and/or Trp10b encoding mRNA.

It has been found that carcinoma cells of placenta (chorion carcinoma), lung and prostate express Trp10 transcripts as well as Trp10 antisense transcripts and transcripts being in part complementary to Trp10 antisense transcripts. Accordingly, the present invention also provides a method for diagnosing a melanoma, chorion carcinoma, cancer of the lung and of the prostate in a tissue of a subject, comprising contacting a sample with a reagent which detects Trp10a and/or Trp10b antisense RNA.

When the target is mRNA (or antisense RNA), the reagent is typically a nucleic acid probe or a primer for PCR. The person skilled in the art is in a position to design suitable nucleic acids probes based on the information as regards the nucleotide sequence of Trp8a, Trp8b, Trp10a or Trp10b as depicted in figure 7, 8a, 10 and 11, respectively, or tables 1 and 2, above. When the target is the protein, the reagent is typically an antibody probe. The term "antibody", preferably, relates to antibodies which consist essentially of pooled monoclonal antibodies with different epitopic specifities, as well as distinct monoclonal antibody preparations, Monoclonal antibodies are made from an antigen containing fragments of the proteins of the invention by methods well known to those skilled in the art (see, e.g., Köhler et al., Nature 256 (1975), 495). As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, Fab and F(ab') 2 fragments) which are capable of specifically binding to protein. Fab and f(ab')2 fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding than an intact antibody. (Wahl et al., J. Nucl. Med. 24: 316-325 (1983)). Thus, these fragments are preferred, as well as the products of a FAB or other immunoglobulin expression library. Moreover, antibodies of the present invention include chimerical, single chain, and humanized antibodies. The target cellular component, i.e. Trp8a, Trp8b, Trp10a and/or Trp10b or Trp8a, Trp8b, Trp10a and/or Trp10b encoding mRNA or Trp10a/b antisense transcripts, e.g., in biological fluids or tissues, may be detected directly in situ, e.g. by in situ hybridization (e.g., according to the examples, below) or it may be isolated from other cell components by common methods known to those skilled in the art before contacting with a probe. Detection methods include Northern blot analysis, RNase protection, in situ methods, e.g. in situ hybridization, in vitro amplification methods (PCR, LCR, QRNA replicase or RNA-transcription/amplification (TAS, 3SR), reverse dot blot disclosed in EP-B1 O 237 362)), immunoassays, Western blot and other detection assays that are known to those skilled in the art.

Products obtained by in vitro amplification can be detected according to established methods, e.g. by separating the products on agarose gels and by subsequent staining with ethidium bromide. Alternatively, the amplified products can be detected by using labeled primers for amplification or labeled dNTPs.

The probes can be detectable labeled, for example, with a radioisotope, a bioluminescent, compound, a chemiluminescent compound, a fluorescent compound, a metal chelate, or an enzyme.

Expression of Trp8a, Trp10a and Trp10b, respectively, in tissues can be studied with classical immunohistological methods (Jalkanen et al., J. Cell. Biol. 101 (1985), 976-985; Jalkanen et al., J. Cell. Biol. 105 (1987), 3087-3096; Sobol et al. Clin. Immunpathol. 24 (1982), 139-144; Sobol et al., Cancer 65 (1985), 2005-2010). Other antibody based methods useful for detecting protein gene expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA). Suitable antibody assay labels are known in the art and include enzyme labels, such as, glucose oxidase, and radioisotopes, such as iodine (125 I, 121 I), carbon (14C), sulfur (35S), tritium (3H), indium (112 In), and technetium rhodamine, and biotin. In addition to assaying Trp8a, Trp8b, Trp 10a or Trp10b levels in a biological sample, the protein can also be detected in vivo by imaging. Antibody labels or markers for in vivo imaging of protein include those detectable by Xradiography, NMR or ESR. For X-radiography, suitable labels include radioisotopes such as barium or cesium, which emit detectable radiation but are not overtly harmful to the subject. Suitable markers for NMR and ESR include those with a detectable characteristic spin, such as deuterium, which may be incorporated into the antibody by labeling of nutrients for the relevant hybridoma. A protein-specific antibody or antibody fragment which has been labeled with an appropriate detectable imaging moiety, such as a radioisotope (for example, <sup>131</sup>L, <sup>112</sup>In, <sup>99</sup>mTc), a radio-opaque substance, or a material detectable by nuclear magnetic resonance, is introduced (for example, parenterally, subcutaneously, or intraperitoneally) into the mammal. It will be understood in the art that the size of the subject and the imaging system used will determine the quantity of imaging moiety needed to produce diagnostic images. In the case of a radioisotope moiety, for a human subject, the quantity of radioactivity injected will normally range from about 5 to 20 millicuries of <sup>99</sup>mTc. The labeled antibody or antibody fragment will then preferentially accumulate at he location of cells which contain the specific protein. In

vivo tumor imaging is described in S.W. Burchiel et al., "Immunopharmacokinetics of Radiolabeled Antibodies and Their Fragments". (Chapter 13 in Tumor Imaging: The Radiochemical Detection of Cancer, S.W. Burchiel and B.A. Rhodes, eds., Masson Publishing Inc. (1982)).

The marker Trp8a and Trp8b is also useful for prognosis, for monitoring the progression of the tumor and the diagnostic evaluation of the degree of malignancy of a prostate tumor (grading and staging), e.g. by using in situ hybridization: In a primary carcinoma Trp8 is expressed in about 2 to 10% of carcinoma cells, in a rezidive carcinoma in about 10 to 60% of cells and in metastases in about 60 to 90% of cells.

The present invention also relates to a method for diagnosing endometrial cancer (cancer of the uterus) which comprises contacting a target sample suspected to contain the protein Trp8a and/or Trp8b or the Trp8a and/or Trp8b encoding mRNA with a reagent which reacts with Trp8a and/or Trp8b or the encoding mRNA and detecting Trp8a and/or Trp8b encoding mRNA. As regards particular embodiments of this method reference is made to the particular embodiments of the method of diagnosing a prostate cancer outlined above.

For evaluating whether the concentration of Trp8a, Trp8b, Trp10a or Trp10b or the concentration of Trp8a, Trp8b, Trp10a or Trp10b encoding mRNA is normal or increased, thus indicative for the presence of a malignant tumor, the measured concentration is compared with the concentration in a normal tissue, preferably by using the ratio of Trp8a:Trp9, Trp8b:Trp9 or Trp10(a or b)/Trp9 for quantification.

Since the prostate carcinoma forms its own basement membrane when growing invasively, it can be concluded that only cells expressing Trp8 and Trp10 are involved in this phenomenon. Thus, it can be concluded that by inhibiting the expression and/or activity of these proteins an effective therapy of cancers like PCA is provided.

Thus, the present invention also relates to a pharmaceutical composition containing a reagent which decreases or inhibits Trp8a, Trp8b, Trp10a and/or Trp10b expression or the activity of Trp8a, Trp8b, Trp10a and/or Trp10b, and a method for preventing, treating, or ameliorating a prostate tumor, endometrial cancer (uterine carcinoma) tumor, a chorion carcinoma, cancer of the lung or melanoma, which comprises administering to a mammalian subject a

therapeutically effective amount of a reagent which decreases or inhibits Trp8a, Trp8b, Trp10a and/or Trp10b expression or the activity of Trp8a, Trp8b, Trp10a and/or Trp10b. Examples of such reagents are the above described antisense RNAs, ribozymes or inhibitors, e.g. specific antibodies. Furthermore, peptides, which inhibit or modulate the biological function of Trp8a, Trp8b, Trp9, Trp10a and/or Trp10b may be useful as therapeutical reagents. For example, these peptides can be obtained by screening combina torial phage display libraries (Cosmix, Braunschweig, Germany) as described by Rottgen, P. and Collins, J. (Gene (1995) 164 (2): 243-250). Furthermore, antigenic epitopes of the Trp8 and Trp10 proteins can be identified by the expression of recombinant Trp8 and Trp10 epitope libraries in E. coli (Marquart, A. & Flockerzi, V., FEBS Lett. 407 (1997), 137-140; Trost, C., et al., FEBS Lett. 451 (1999) 257-263 and the consecutive screening of these libraries with serum of patients with cancer of the prostate or of the endometrium. Those Trp8 and Trp10 epitopes which are immunogenic and which lead to the formation of antibodies in the serum of the patients can be then be used as Trp8 or Trp10 derived peptide vaccines for immune inventions against cancer cells which express Trp8 or Trp10. Alternatively to the E. coli expression system, Trp8 or Trp10 or epitopes of Trp8 and Trp10 can be expressed in mammalian cell lines such as human embryonic kidney (Hek 293) cells (American Type Culture Collection, ATCC CRL 1573).

Finally, compounds useful for therapy of the above described diseases comprise compounds which act as antagonists or agonists on the ion channels Trp8, Trp9 and Trp10. It could be shown that Trp8 is a highly calcium selective ion channel which in the presence of monovalent (namely sodium) and divalent ions (namely calcium) is only permeable for calcium ions (see Example 4, below, and Figures 3A, C, E). Under physiological conditions, Trp8 is a calcium selective channel exhibiting large inward currents. This very large conductance of Trp8 channels (as wells as Trp9 and Trp10a/b channels) is useful to establish systems for screening pharmacological compounds interacting with Trp-channels including high throughput screening systems. Useful high throughput screening systems are well known to the person skilled in the art and include, e.g., the use of cell lines stably or transiently transfected with DNA sequences encoding Trp8, Trp9 and Trp10 channels in assays to detect calcium signaling in biological systems. Such systems include assays based on Ca-sensitive dyes such as aequorin, apoaequorin, Fura-2, Fluo-3 and Indo-1.

Accordingly, the present invention also relates to a method for identifying compounds which act as agonists or antagonists on the ion channels Trp8, Trp9 and/or Trp10, said method comprising contacting a test compound with the ion channel Trp8, Trp9 and/or Trp10, preferably by using a system based on cells stably or transiently transfected with DNA sequences encoding Trp8, Trp9 and/or Trp10, and determining whether said test compound affects the calcium uptake.

For administration the above described reagents are preferably combined with suitable pharmaceutical carriers. Examples of suitable pharmaceutical carriers are well known in the art and include phosphate buffered saline solutions, water, emulsions, such as oil/water emulsions, various types of wetting agents, sterile solutions etc. Such carriers can be formulated by conventional methods and can be administered to the subject at a suitable dose. Administration of the suitable compositions may be effected by different ways, e.g. by intravenous, intraperetoneal subcutaneous, intramuscular, topical or intradermal administration. The route of administration, of course, depends on the nature of the tumor and the kind of compound contained in the pharmaceutical composition. The dosage regimen will be determined by the attending physician and other clinical factors. As is well known in the medical arts, dosages for any one patient depends on many factors, including the patient's size, body surface area, age, sex, the particular compound to be administered, time and route of administration, the kind and stage of the tumor, general health and other drugs being administered concurrently.

The delivery of the antisense RNAs or ribozymes of the invention can be achieved by direct application or, preferably, by using a recombinant expression vector such as a chimeric virus containing these compounds or a colloidal dispersion system. By delivering these nucleic acids to the desired target, the intracellular expression of Trp8a, Trp8b, Trp10a and/or Trp10b and, thus, the level of Trp8a, Trp8b, Trp10a and/or Trp10b can be decreased resulting in the inhibition of the negative effects of Trp8a, Trp8b, Trp10a and/or Trp10b, e.g. as regards the metastasis formation of PCA.

Direct application to the target site can be performed, e.g., by ballistic delivery, as a colloidal dispersion system or by catheter to a site in artery. The colloidal dispersion systems which can be used for delivery of the above nucleic, acids include macromolecule complexes, nanocapsules, microspheres, beads and lipid-based systems including oil-in-water emulsions

(mixed), micelles, liposomes and lipoplexes, The preferred colloidal system is a liposome. The composition of the liposome is usually a combination of phospholipids and steroids, especially cholesterol. The skilled person is in a position to select such liposomes which are suitable for the delivery of the desired nucleic acid molecule. Organ-specific or cell-specific liposomes can be used in order to achieve delivery only to the desired tumor. The targeting of liposomes can be carried out by the person skilled in the art by applying commonly known methods. This targeting includes passive targeting (utilizing the natural tendency of the liposomes to distribute to cells of the RES in organs which contain sinusoidal capillaries) or active targeting (for example by coupling the liposome to a specific ligand, e.g., an antibody, a receptor, sugar, glycolipid, protein etc., by well known methods). In the present invention monoclonal antibodies are preferably used to target liposomes to specific tumors via specific cell-surface ligands.

Preferred recombinant vectors useful for gene therapy are viral vectors, e.g. adenovirus, herpes virus, vaccinia, or, more preferably, an RNA virus such as a Retrovirus. Even more preferably, the retroviral vector is a derivative of a murine or avian retrovirus. Examples of such retroviral vectors which can be used in the present invention are: Moloney murine leukemia virus (MoMuLV), Harvey murine sarcoma virus (HaMuSV), murine mammary tumor virus (MuMTV) and Rous sarcoma virus (RSV). Most preferably, a non-human primate retroviral vector is employed, such as the gibbon ape leukemia virus (GaLV), providing a broader host range compared to murine vectors. Since recombinant retroviruses are defective, assistance is required in order to produce infectious particles. Such assistance can be provided, e.g., by using helper cell lines that contain plasmids encoding all of the structural genes of the retrovirus under the control of regulatory sequences within the LTR. Suitable helper cell lines are well known to those skilled in the art. Said vectors can additionally contain a gene encoding a selectable marker so that the transduced cells can be identified. Moreover, the retroviral vectors can be modified in such a way that they become target specific. This can be achieved, e.g., by inserting a polynucleotide encoding a sugar, a glycolipid, or a protein, preferably an antibody. Those skilled in the art know additional methods for generating target specific vectors. Further suitable vectors and methods for in vitro- or in vivo-gene therapy are described in the literature and are known to the persons skilled in the art; see, e.g., WO 94/29469 or WO 97/00957.

In order to achieve expression only in the target organ, i.e. tumor to be treated, the nucleic acids encoding, e.g. an antisense RNA or ribozyme can also be operably linked to a tissue specific promoter and used for gene therapy. Such promoters are well known to those skilled in the art (see e.g. Zimmermann et al., (1994) Neuron 12, 11-24; Vidal et al.; (1990) EMBO J. 9, 833-840; Mayford et al., (1995), Cell 81, 891-904; Pinkert et al., (1987) Genes & Dev. 1, 268-76).

For use in the diagnostic research discussed above, kits are also provided by the present invention. Such kits are useful for the detection of a target cellular component, which is Trp8a, Trp8b, Trp10a and/or Trp10b or, alternatively, Trp8a, Trp8b, Trp10a and/or Trp10b encoding mRNA or Trp10a/b antisense transcripts, wherein the presence or an increased concentration of Trp8a, Trp8b, Trp10a and/or Trp10b or, alternatively, Trp8a, Trp8b, Trp10a and/or Trp10b encoding mRNA or Trp10a/b antisense transcripts is indicative for a prostate tumor, endometrial cancer, melanoma, chorion carcinoma or cancer of the lung, said kit comprising a probe for detection of Trp8a, Trp8b, Trp9, Trp10a and/or Trp10b or, alternatively, Trp8a, Trp8b, Trp9, Trp10a and/or Trp10b encoding mRNA or Trp10a/b antisense transcripts. The probe can be detectably labeled. Such probe may be a specific antibody or specific oligonucleotide. In a preferred embodiment, said kit contains an anti-Trp8a-, anti-Trp8b-, anti-Trp10a-and/or anti-Trp10b-antibody and allows said diagnosis, e.g., by ELISA and contains the antibody bound to a solid support, for example, a polystyrene microtiter dish or nitrocellulose paper, using techniques known in the art. Alternatively, said kits are based on a RIA and contain said antibody marked with a radioactive isotope. In a preferred embodiment of the kit of the invention the antibody is labeled with enzymes, fluorescent compounds, luminescent compounds, ferromagnetic probes or radioactive compounds. The kit of the invention may comprise one or more containers filled with, for example, one or more probes of the invention. Associated with container (s) of the kit can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, us or sale for human administration.

#### **EXAMPLES**

The following Examples are intended to illustrate, but not to limit the invention. While such Examples are typical of those that might be used, other methods known to those skilled in the art may alternatively be utilized.

## **Example 1: Materials and Methods**

## (A) Isolation of cDNA clones and Northern blot analysis

Total RNA was isolated from human placenta an prostate using standard techniques. Isolation of mRNA was performed with poly (A)<sup>†</sup>RNA - spin columns (New England Biolabs, Beverly, USA) according to the instructions of the manufacturer. Poly (a) <sup>†</sup>RNA was reverse transcribed using the cDNA choice system (Gibco-BRL, Rockville, USA) and subcloned in λ-Zap phages (Stratagene, La Jolla, USA). An human expressed sequence tag (GenBank accession number 1404042) was used to screen an oligo d(T) primed human placenta cDNA library. Several cDNA clones were identified and isolated. Additional cDNA clones were isolated from two specifically primed cDNA libraries using primers 5'-gca tag gaa ggg aca ggt gg-3' and 5'-gag agt cga ggt cag tgg tcc-3'.

cDNA clones were sequenced using a thermocycler (PE Applied Biosystems, USA) and Thermo Sequenase (Amersham Pharmacia Biotech Europe, Freiburg, Germany). DNA sequences were analyzed with an automated sequencer (Licor, Linccoln, USA).

For Northern blot analysis 5  $\mu$ g human poly (A)<sup>+</sup> RNA from human placenta or prostate were separated by electrophoresis on 0.8 % agarose gels. Poly (A)<sup>+</sup> RNA was transferred to Hybond N nylon membranes (Amersham Pharmacia Biotech Europe, Freiburg, Germany). The membranes were hyridized in the presence of 50 % formamide at 42°C over night. DNA probes were labelled using  $[\alpha^{32}P]dCTP$  and the "ready prime, labelling kit (Amersham Pharmacia Biotech Europe, Freiburg, Germany). Commercial Northern blots were hybridized according to the distributors instructions (Clontech, Paolo Alto, USA).

# (B) Construction of expression plasmids and transfection of HEK 293 cells

Lipofections were carried out with the recombinant dicistronic eucaryotic expression plasmid pdiTRP8 containing the cDNA of Trp8b under the control of the chicken B-actin promotor followed by an internal ribosome entry side (IRES) and the cDNA of the green fluorescent protein (GFP). To obtain pdiTRP8 carrying the entire protein coding regions of TRP8b and

the GFP (Prasher, D.C. et al. (1992), Gene 111, 229-233), the 5'and 3'-untranslated sequences of the TRP8b cDNA were removed, the consensus sequence for initiation of translation in vertebrates (Kozak, M. (1987) Nucleic Acids Research15, 8125-8148) was introduced immediately 5'of the translation initiation codon and the resulting cDNA was subcloned into the pCAGGS vector (Niwa, H., Yamamura, K. and Miyazaki, J (1991), Gene 8, 193-199) downstream of the chicken β-actin promotor. The IRES derived from encephalmyocarditis virus (Kim, D.G., Kang, H.M., Jang, S.K. and Shin H.S. (1992) Mol.Cell.Biol. 12, 3636-3643) followed by the GFP cDNA containing a Ser65Thr mutation (Heim, R., Cubitt, A.B., Tsien, R.Y. (1995) Nature 373, 663-664) was then cloned 3' to the TRP8b cDNA. The IRES sequence allows the simultaneous translation of TRP8b and GFP from one transcript. Thus, transfected cells can be detected unequivocally by the development of green fluorescence.

For monitoring of the intracellular Ca<sup>2+</sup> concentration human embryonic kidney (HEK 293) cells were cotransfected with the pcDNA3-TRP8b vector and the pcDNA3-GFPvector in a molar ratio of 4:1 in the presence of lipofectamine (Quiagen, Hilden, Germany). To obtain pcDNA3-TRP8b the entire protein coding region of TRP8b including the consensus sequence for initiation of translation in vertebrates (Kozak, M. (1987) Nucleic Acids Research15, 8125-8148) was subcloned into the pcDNA3 vector (Invitrogen, Groningen, Netherlands). Calcium monitoring and patch clamp experiments were carried out two days and one day after transfection, respectively.

## (C) Chromosomal localization of the Trp8 gene

The chromosomal localization of the human TRP8 gene was performed using NIGMS somatic hybrid mapping panel No.2 (Coriell Institute, Camden, NJ, USA) previously described (Drwinga, H.L., Toji, L.H., Kim, C.H., Greene, A.E., Mulivor, R.A. (1993) Genomics 16, 311-314; Dubois, B.L. and Naylor, S.L. (1993) Genomics 16, 315-319).

(D) In Vitro Translation, glutathione - sepharose and calmodulin agarose binding assay N- and C-terminal Trp8-fragments were subcloned into the pGEX-4T2 vector (Amersham Pharmacia Europe, Freiburg, Germany) resulting in glutathione-S-transferase (GST)-Trp8 fusion constructs (Fig. 4). The GST-TRP8-fusion proteins were expressed in E. coli BL 21 cells and purified using glutathione - sepharose beads (Amersham Pharmacia Biotech Europe, Freiburg, Germany).

In vitro translation of human Trp8 cDNA and Xenopus laevis calmodulin cDNA (Davis, T.N. and Thorner, J. Proc.Natl.Acad.Sci. USA 86, 7909-7913.) was performed in the presence of <sup>35</sup>S-methionine using the TNT coupled transcription/translation kit (Promega, Madison, USA). Translation products were purified by gel fliltration (Sephadex G50, Amersham Pharmacia Biotech Europe, Freiburg, Germany) and equal amounts of <sup>35</sup>S labeled probes were incubated for 2 h with glutathione beads bound to GST - Trp8 or calmodulin - agarose (Calbiochem) in 50 mM Tris-HCl, pH 7.4, 0.1 % Triton X-100, 150 mM NaCl in the presence of 1 mM Ca<sup>2+</sup> or 2 mM EGTA. After three washes, bound proteins were eluted with SDS sample buffer, fractionated by SDS-PAGE and <sup>35</sup>S labeled proteins were detected using a Phosphor Imager (Fujifilm, Tokyo, Japan).

# (E) Calcium measurements

The intracellular Ca<sup>2+</sup> concentration ([Ca<sup>2+</sup>]<sub>i</sub>) was determined by dual wavelength fura-2 fluorescence ratio measurements (Tsien, R.Y. (1988) Trends Neurosci. 11, 419-424) using a digital imaging system (T.I.L.L. Photonics, Planegg, Germany). HEK cells were grown in minimal essential medium in the presence of 10 % fetal calf serum and cotransfected with the pcDNA3-TRP8b vector and the pcDNA3-GFPvector as described above (B). Transfected cells were detected by development of green fluorescence. The cells were loaded with 4μM fura-2/AM (Molecular Probes, Oregon, USA) for one hour. After loading the cells were rinsed 3 times with buffer B1 (10 mM Hepes, 115 mM NaCl, 2 mM MgCl<sub>2</sub>, 5mM KCl, pH 7.4) and the [Ca<sup>2+</sup>]<sub>i</sub> was calculated from the fluorescence ratios obtained at 340 and 380 nm excitation wavelengths as described (Garcia, D.E., Cavalié, A. and Lux, H.D. (1994) J. Neurosci 14, 545-553).

#### (F) Electrophysiological recordings

HEK cells were transfected with the eucaryotic expression plasmid pdiTRP8 described in (B) and electrophysiological recordings were carried out one day after transfection. Single cells were voltage clamped in the whole cell mode of the patch clamp technique as described (Hamill, O.P., Marty, A., Neher, E., Sakmann, B. and Sigworth, F.J. (1981) Pflügers Arch. 391, 85-100; Philipp, S., Cavalié, A., Freichel, M., Wissenbach, U., Zimmer, S., Trost, C., Marquart, A., Murakami, M. and Flockerzi, V. (1996) EMBO J. 6166-6171). The pipette solution contained contained (mM): 140 aspartic acid, 10 EGTA, 10 NaCl, 1 MgCl2, 10 Hepes (pH 7.2 with CsOH) or 125 CsCl, 10 EGTA, 4 CaCl<sub>2</sub> 10 Hepes (pH 7,2 with CsOH). The bath solution contained (mM): 100 NaCl, 10 CsCl, 2 MgCl<sub>2</sub>, 50 mannitol, 10 glucose, 20

Hepes (pH 7,4 with CsOH) and 2 CaCl<sub>2</sub>, or no added CaCl<sub>2</sub> (-Ca<sup>2+</sup> solution). Divalent free bath solution contained (mM): 110 N-methyl-D-glucamine (NMDG). Whole cell currents were recorded during 100 msec voltage ramps from -100 to +100 mV at varying holding potentials.

## (G) In Situ Hybridization

In situ hybridizations were carried out using formalin fixed tissue slices of 6 - 8 µM thickness. The slices were hydrated and incubated in the presence of PBS buffer including 10 µg / ml proteinase K (Roche Diagnostics, Mannheim, Germany) for 0.5 h. The slices were hybridized at 37°C using biotinylated deoxy-oligonucleotides (0.5 pmol / µl) in the presence of 33 % formamide for 12 h. Furthermore the slices were several times rinsed with 2 x SSC and incubated at 25°C for 0.5 h with avidin / biotinylated horse raddish peroxidase complex (ABC, DAKO, Santa Barbara, USA). After several washes with PBS buffer the slices were incubated in the presence of biotinylated tyramid and peroxide (0.15 % w/v) for 10 min, rinsed with PBS buffer and additionally incubated with ABC complex for 0.5 h. The slices were washed with PBS buffer and incubated in the presence of DAB solution (diaminobenzidine (50µg / ml), 50 mM Tris/EDTA buffer pH 8.4, 0.15 % H<sub>2</sub>O<sub>2</sub> in N,N dimethyl-formamide; Merck, Darmstadt, Germany), The detection was stopped after 4 minutes by incubating the slides in water. Tyramid was biotinylated by incubating NHS-LC Biotin (sulfosuccinimidyl-6-(biotinimid)-hexanoat), 2.5 mg/ml; Pierce, Rockford, USA) and tyramin-HCl (0.75 mg / ml, Sigma) in 25 mM borate buffer pH 8.5 for 12 h. The tyramid solution was diluted 1 - 5: 1000 in PBS buffer.

(H) GenBank accession numbers: TRP8a, Aj243500; TRP8b Aj243501

## **Example 2: Expression of TRP8 transcripts**

In search of proteins distantly related to the TRP family of ion channels, an human expressed sequence tag (EST, GenBank accession number 1404042) was identified in the GenBank database using BLAST programms (at the National Center for Biotechnology Information (NCBI); Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J.J. (1990) Mol. Biol. 5, 403-410) being slightly homologous to the VR1 gene. Several human placenta cDNA libraries were constructed and screeened with this EST DNA as probe. Several full length

cDNA clones were identified and isolated. The full length cDNA clones encoded two putative proteins differing in three amino acids and were termed Trp8a and Trp8b (Fig. 1c, 2a, 7 and 8A). This finding was reproduced by isolating cDNA clones from two cDNA libraries constructed from two individual placentas. The derived protein sequence(s) comprises six transmembrane domains, a characteristic overall feature of trp channels and related proteins (Fig.: 1b). The sequence is closely related to the meanwhile published calcium uptake transport protein 1 (CaT1), isolated from rat intestine (Peng, J.B., Chen, X.Z., Berger, U.V., Vassilev, P.M., Tsukaguchi, H., Brown, E.M. and Hediger M.A. (1999) J Biol Chem. 6;274, 22739-22746) and to the epithelial calcium uptake channel (ECaC) isolated from rabbit kidney (Hoenderop, J.G., van der Kemp, A.W., Hartog, A., van de Graaf, S.F., van Os, C.H., Willems, P.H. and Bindels, R.J. (1999) J Biol Chem. 26;274, 8375-8378). Expression of Trp8a/b transcripts are detectable in human placenta, pancreas and prostate (Fig.: 5) and the size of the Northern signal (3.0 kb) corresponds with the size of the isolated full length cDNAs. In addition, a shorter transcript of 1.8 kb, probably a splice variant, is detectable in human testis. The Trp8 mRNA is not expressed in small intestine or colon (Fig.: 5) implicating that Trp8 is not the human ortholog of the rat CaT1 or rabbit ECaC proteins. To investigate whether there are other related sequences Trp8a/b derived primers (UW241, 5'-TAT GAG GGT TCA GAC TGC-3' and UW242, 5'-CAA AGT AGA TGA GGT TGC-3') were used to amplify a 105 bp fragment from human genomic DNA being 95% identical on the nucleotide level to the Trp8 sequence (data not shown). This indicates the existence of several similar sequences in humans at least at the genomic level.

# Example 3: Two variants of the Trp8 protein (Trp8a and Trp8b) arise by polymorphism

Two variants of the Trp8 cDNA were isolated from human placenta (Fig.: 2A, 7 and 8A) which encoded two proteins which differ in three amino acids and were termed Trp8a and Trp8b. Trp8a/b specific primers were designed to amplify a DNA fragment of 458 bp of the Trp8 gene from genomic DNA isolated from human T-lymphocytes (primer pair: UW243, 5'-CAC CAT GTG CTG CAT CTA CC-3' and UW244, 5'-CAA TGA CAG TCA CCA GCT CC-3'). The amplification product contains a part of the sequence where the derived protein sequence of Trp8a comprises the amino acid valine and the Trp8b sequence methionine as well as a silent base pair exchange (g versus a) and an intron of 303bp (Fig.: 2.A, B). Both variants of the Trp8 genes (a,b) were amplified from genomic DNA in equal amounts indicating the existence of both variants in the human genome and therefore being not the

result of RNA editing (Fig.: 2B). The Trp8a gene can be distinguished from the Trp8b gene by cutting the genomic fragment of 458bp with the restriction enzyme Bsp1286I (Fig. 2B). Using human genomic DNA isolated from blood of twelve human subjects as template, the 458bp fragment was amplified and restricted with BSP1286I. In eleven of the tested subjects only the Trp8b gene is detectable, while one subject (7) contains Trp8a and Trp8b genes (Fig.: 2D). These implicates that the two Trp8 variants arise by polymorphism and do not represent individual genes. Using Trp8 specific primers and chromosomal DNA as template, the Trp8 locus is detectable on chromosome 7 (Fig.: 2C).

## Example 4: Trp8b is a calcium permeable channel

The protein coding sequence of the Trp8b cDNA was subcloned into pcDNA3 vector (Invitrogen, Groningen, Netherlands) under the control of the cytomegalovirus promotor (CMV). Human embryonic kidney (HEK 293) cells were cotransfected with the Trp8b pcDNA3 construct (pcDNA3-Trp8b vector) and the pcDNA3-GFPvector encoding the green fluorescent protein (GFP) in 4:1 ratio. The Trp8b cDNA and the cDNA of the reporter, GFP, was transiently expressed in human embryonic kidney (HEK 293) cells. The intracellular Ca<sup>2+</sup> concentration ([Ca<sup>2+</sup>]<sub>i</sub>) and changes of [Ca<sup>2+</sup>]<sub>i</sub> were determined by dual wavelength fura-2 fluorescence ratio measurements (Fig.: 3F) in cotransfected cells which were identified by the green fluorescence of the reporter gene GFP.

Dual wavelength fura-2 fluorescence ratio measurement is a standard procedure (e.g. in: An introduction of Molecular Neurobiology (ed. Hall, Z.W.)Sinauer Associates, Sunderland, USA (1992)) using fura-2, which is a fluorescent Ca<sup>2+</sup> sensitive dye and which was designed by R.Y.Tsien (e.g. Trends Neurosci. 11, 419-424 (1988) based upon the structure of EGTA. Its fluorescence emission spectrum is altered by binding to Ca<sup>2+</sup> in the physiological concentration range. In the absence of Ca<sup>2+</sup>, fura-2 fluoresces most strongly at an excitation wavelength of 385 nm; when it binds Ca<sup>2+</sup>, the most effective excitation wavelength shifts to 345 nm. This property is used to measure local Ca<sup>2+</sup> concentrations within cells. Cells can be loaded with fura-2 esters (e.g. fura-2AM) that diffuse across cell membranes and are hydrolyzed to active fura-2 by cytosolic esterases.

In the presence of 1mM Ca<sup>2+</sup>, Trp8 expressing cells typically contained more than 300 nM cytosolic Ca<sup>2+</sup>, while non transfected controls contained less than 100 nM Ca<sup>2+</sup> ions (Fig. 3F).

When Trp8b transfected cells were incubated without extracellular Ca<sup>2+</sup>, the intracellular Ca<sup>2+</sup> concentration ([Ca<sup>2+</sup>]<sub>i</sub>) decreased to levels comparable to non transfected cells. Readdition of 1mM Ca<sup>2+</sup> to the bath resulted in significant increase of the cytosolic [Ca<sup>2+</sup>] in Trp8b transfected cells, but not in controls (Fig.: 3F). After readdition of Ca<sup>2+</sup> ions to the bath solution, the cytosolic Ca<sup>2+</sup> concentration remains on a high steady state level in the Trp8b transfected cells.

#### Example 5: Trp8 expressing cells show calcium selective inward currents

To characterize in detail the electrophysiological properties of TRP8, TRP8 and GFP were coexpressed in HEK293 cells using the dicistronic expression vector pdiTRP8 and measured currents using the patch clamp technique in the whole cell mode (Hamill, O.P., Marty, A., Neher, E., Sakmann, B. and Sigworth, F.J. (1981) Pflugers Arch., 391, 85-100).

The eucaryotic expression plasmid pdiTRP8 contains the cDNA of Trp8b under the control of the chicken β-actin promotor followed by an internal ribosome entry side (IRES) and the cDNA of the green fluorescent protein (GFP). To obtain pdiTRP8 carrying the entire protein coding regions of TRP8b and the GFP (Prasher, D.C. et al. (1992), Gene 111, 229-233), the 5'and 3'-untranslated sequences of the TRP8b cDNA were removed, the consensus sequence for initiation of translation in vertebrates (Kozak, M. (1987) Nucleic Acids Research15, 8125-8148) was introduced immediately 5'of the translation initiation codon and the resulting cDNA was subcloned into the pCAGGS vector (Niwa, H., Yamamura, K. and Miyazaki, J (1991), Gene 8, 193-199) downstream of the chicken β-actin promotor. The IRES derived from encephalmyocarditis virus (Kim, D.G., Kang, H.M., Jang, S.K. and Shin H.S. (1992) Mol.Cell.Biol. 12, 3636-3643) followed by the GFP cDNA containing a Ser65Thr mutation (Heim, R., Cubitt, A.B., Tsien, R.Y. (1995) Nature 373, 663-664) was then cloned 3' to the TRP8b cDNA. The IRES sequence allows the simultaneous translation of TRP8b and GFP from one transcript. Thus, transfected cells can be detected unequivocally by the development of green fluorescence.

In the presence of 2 mM external calcium, Trp8b transfected HEK cells show inwardly rectifying currents, the size of which depends on the level of intracellular calcium and the electrochemical driving force. The resting membrane potential was held either at -40 mV, or, to lower the driving force for calcium influx in between pulses, at + 70 mV. Current traces

were recorded in response to voltage ramps from -100 to +100 mV, that were applied every second. To monitor inward and outward currents over time, we analyzed the current size at -80 and + 80 mV of the ramps. Figure 3A shows a representative trace of the current at - 80 mV over time. Both at a holding potential of -40 mV or at +70 mV, the currents are significantly larger than in cells transfected with only the GFP containing vector (Fig.: 3E). Interestingly, after changing to a positive holding potential, current size in Trp8 transfected cells slowly increases and reaches steady state after approximately 70 seconds (Fig.: 3A). To determine the selectivity of the induced currents, we then perfused the cells with solutions that either contain no sodium, no added Ca<sup>2+</sup> (Fig. 3A, C) or a sodium containing, but divalent ion free bath solution. To control for the effect of the solution change alone, we also perfused with normal bath (see puff in Fig. 3A). While removal of external Ca<sup>2+</sup> completely abolishes the trp 8 induced currents - the remaining current being identical in size and shape to the control (Fig.: 3A, C, E), removal of external sodium has no effect (Fig.: 3E). An important hallmark of calcium selective channels (e.g. Vennekens, R., Hoenderop, G.J., Prenen, J., Stuiover, M., Willems, PHGM, Droogmans, G., Nilius, B.and Bindels, R.J.M (1999) J. Biol. Chem. 275, 3963-3969), is their ability to conduct sodium only if all external divalent ions, namely Ca<sup>2+</sup> and magnesium are removed. To test whether the trp 8 channel conforms with this phenomenon normal bath solution was switched to a solution containing only sodium and 1 mM EGTA. As can be seen in Figure 3B and D, Trp8 transfected cells can now conduct very large sodium currents. Interestingly, immediately after the solution change, the currents first become smaller before increasing rapidly, indicating that the pore may initially still be blocked by calcium a phenomenon usually called anomalous mole fraction behaviour (Warnat, J., Philipp, S., Zimmer, S., Flockerzi, V., and Cavalié A.(1999) J.Physiol. (Lond) 518, 631-638). The measured outward currents of Trp8 transfected cells in normal bath solution are not significantly different from non-transfected control cells or cells which only express the reporter gene GFP. As the removal of external Ca<sup>2+</sup> abolishes the Trp8 specific current, the remaining current was subtracted from the current before the solution change to obtain the uncontaminated Trp8 conductance (see inset in Fig.: 3C). As expected from the given ionic conditions (high EGTA inside, 2 mM Ca2+ outside), the current-voltage relationship now shows prominent inward rectification with little to no outward current.

Both the time course of the development of Trp8 currents and the size of the currents depend on the frequency of stimulation (data not shown), the internal and external Ca<sup>2+</sup> concentration

and the resting membrane potential, suggesting that Trp8 calcium conductance is intrically regulated by a Ca<sup>2+</sup> mediated feedback mechanisms.

# Example 6: Ca<sup>2+</sup> / calmodulin binds to the C-terminus of the Trp8 protein

To test whether calmodulin, a prime mediator of calcium regulated feedback, is involved, first it was investigated biochemically whether Trp8 protein can bind calmodulin. Trp8 cDNA was in vitro translated in the presence of <sup>35</sup>S-methionine and the product incubated with calmodulin-agarose beads. After several washes either in the presence or abscence of Ca<sup>2+</sup>, the beads were incubated in Laemmli buffer and subjected to SDS-polyacrylamide gel electrophoresis. In the presence of Ca<sup>2+</sup> (1mM), but not in the absence of Ca<sup>2+</sup>, Trp8 protein binds to calmodulin (Fig.: 4B).

To narrow down the binding site, two approaches were undertaken: Firstly, GST-TRP8 fusion proteins of various intracellular domains of Trp8 were constructed, expressed in E. coli and bound to gluthathione sepharose beads. These beads were then incubated with in vitro translated <sup>35</sup>S- labeled calmodulin, washed and subjected to gel electrophoresis. Secondly, truncated versions of in vitro translated Trp8 protein were used in the above described binding to calmodulin-agarose. As shown in Figure 4A, and C, fusion proteins of the N-terminal region (N1, N2) of Trp8 did not bind calmodulin, while C-terminal fragments (C1, C2, C3, C4) showed calmodulin binding in the presence of calcium (for localization of fragments within the entire Trp8 protein see Fig. 4C). Accordingly, a truncated version of in vitro translated Trp8, which lacks the C-terminal 32 amino acid residues did not bind to calmodulin-agarose (4B). We have restricted the calmodulin binding site to amino acid residues 691 to 711 of the Trp8 protein. This calmodulin binding site does not resemble the typical conserved IQ - motif of conventional myosins, but has limited sequence homology to the calcium dependent calmodulin binding site 1 of the transient receptor potential like (trpl) protein of Drosophila melanogaster (Warr and Kelly, 1996) with several charged amino acid residues conserved. The sequence of the calmodulin binding site of the Trp8 protein resembles a putative amphipathic α-helical wheel structure with a charged and a hydrophobic site according to a model proposed by Erickson-Vitanen and De Grado (1987, Methods Enzymol. 139, 455-478.).

# Example 7: Expression of Trp8 transcripts in human placenta and pancreas

Several slides from a human placenta of a ten week old abort were used for in situ hybridization experiments. The in situ hybridization experiments revealed expression of Trp8 transcripts in human placenta (Fig.: 5B). Expression was detectable in trophoblasts and syncytiotrophoblasts of the placenta, but not in Langhans cells.

Trp8 transcripts are detectable in human pancreas (Fig.: 5A). Therefore Trp8 probes were hybridized to tissue sections of human pancreas. The pancreatic tissues were removed from patients with pancreas cancer. Trp8 expression is detectable in pancreatic acinar cells, but not in Langerhans islets (Fig.: 5C). No Trp8 expression was found in regions of pancreatic carcinomas (data not shown).

Furthermore, the Trp8 cDNA is not detectable in human colon nor in human kidney by in situ hybridization as well as by Northern analysis (Fig.: 5A, D). The Northern results taken together with the in situ expression data indicate that the Trp8 protein is not the human ortholog of the CaT1 and ECaC channels cloned from rat intestine (Peng, J.B., Chen, X.Z., Berger, U.V., Vassilev, P.M., Tsukaguchi, H., Brown, E.M. and Hediger M.A.(1999) J Biol Chem. 6;274, 22739-22746) and from rabbit kidney (Hoenderop, J.G., van der Kemp, A.W., Hartog, A., van de Graaf, S.F., van Os, C.H., Willems, P.H. and Bindels, R.J. (1999) J Biol Chem. 26;274, 8375-8378), respectively. Trp8 is unlikely to represent the human version of CaT1 as its expression is undetectable in the small intestine and colon tissues where CaT1 is abundantly expressed. If, however, Trp8 is the human version of rat CaT1, a second gene product appears to be required for Ca<sup>2+</sup> uptake in human small intestine and colon attributed to CaT1 in rat small intestine and colon.

# Example 8: Differential expression of Trp8 transcripts in benign and malign tissue of the prostate

The Trp8 transcripts are expressed in human prostate as shown by hybridization of a Trp8 probe to a commercial Northern blot (Clontech, Palo Alto, USA) (Fig.: 5A). Trp8 transcripts were not detectable by Northern blot analysis using pooled mRNA of patients with benign prostatic hyperplasia (BPH) (Fig.: 5A, prostate\*). To examine Trp8 expression on the cellular

level, sections of prostate tissues were hybridized using Trp8 specific cDNA probes (Table 3). Expression of Trp8 transcripts is not detectable in normal prostate (n = 3), benign hyperplasia (BPH, n = 15) or prostatic intraepithelial neoplasia (PIN, n = 9) (Fig.: 6A, C, E). Trp8 transcripts were only detectable in prostate carcinoma (PCA), although with different expression levels. Low expression levels were found in primary carcinomas (2 - 10 % of the carcinoma cells, n = 8) (Fig.: 7B). Much stronger expression was detectable in rezidive carcinoma (10 - 60 %) (Fig.: 7D, n = 6) and metastases of the prostate (60 - 90 %, n = 4) (Fig.: 7F). Thus it has to be concluded that the commercial Northern blot used in Fig.: 5A contains not only normal prostate mRNA as indicated by the distributor. According to the distributors instructions the prostate mRNA used for this Northern blot was collected from 15 human subjects in the range of 14 to 60 years of age. This prostate tissue was not examined by pathologic means. Since Trp8 expression is not detectable in normal or benign prostate, this finding implicates that the mRNA used for this Northern blot was extracted in part from prostatic carcinoma tissue. To summarize, Trp8 expression is only detectable in malign prostate and, thus, the Trp8 cDNA is a marker for prostate carcinoma. The results are summarized in Table 4.

Table 3

Trp8 probes used for in situ hybridization:

Probes (antisense)

- 1.) 5' TCCGCTGCCGGTTGAGATCTTGCC 3'
- 2.) 5' CTTGCTCCATAGGCAGAGAATTAG 3'
- 3.) 5' ATCCTCAGAGCCCCGGGTGTGGAA3'

#### Controls (sense)

- 1.) 5' GGCAAGATCTCAACCGGCAGCGGA 3'
- 2.) 5' CTAATTCTCTGCCTATGGAGCAAG 3'
- 3.) 5' TTCCACACCCGGGGCTCTGAGGAT 3'

Table 4

| Prostate | total | negative | positive |
|----------|-------|----------|----------|
| normal   | 3     | 3        | 0        |
| BPH      | 15    | 15       | 0        |
| PIN      | 9     | 9        | 0        |

carcinoma 18 1 17

#### (B) Differential expression of Trp8 transcripts in benign and malign tissue of the uterus

Moreover it could be shown that Trp8 is expressed in endometrial cancer (also called cancer of the uterus, to be distinguished from uterine sarcoma or cancer of the cervix) whereas no expression was observed in normal uterus tissue. Thus, Trp8 also is a specific marker for the diagnosis of the above cancer (Fig. 12).

### **Example 9: Characterization of Trp9**

The complete protein coding sequence of Trp9 was determined (Fig. 9). Trp 9 transcripts are predominantly expressed in the human prostate and in human colon. As it could be shown by Northern blot analysis, there is no difference of the expression of TRP9 in benigne prostata hyperplasia (BPH, Fig. 13, upper panel left) or prostate carcinoma (Fig. 13, upper panel right). However, Trp9 is useful as a reference marker for prostata carcinoma, i.e. can be used for quantifying the expression level of Trp8. The ratio of the expression of Trp8:Trp9 in patients and healthy individuals is useful for the development of a quantitative assay.

# **Example 10: Characterization of Trp10**

The complete protein coding sequence of TRP10 (a and b) was determined by biocomputing (Fig. 10 and 11). Using a 235 bp fragment of the Trp10 cDNA as probe in Northern blot analysis TRP10 transcripts could only be detected in mRNA isolated from individuals with prostate cancer (Fig. 13, bottom panel) but not in mRNA isolated from benign tissue of the prostate (prostate BPH) nor in mRNA isolated from heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas. The 235 bp cDNA fragment of the Trp10 cDNA was amplified using the primer pair UW248 5'-ACA GCT GCT GGT CTA TTC C-3' and UW249 5'-TAT

GTG CCT TGG TTT GTA CC-3' and prostate cDNA as template. In summary, Trp10a and Trp10b, like TRP8 are also expressed in malignant prostate tissue. So far, its expression could not be observed in any other tissue examined (see above). Thus, Trp 10a and Trp10b are also useful markers which are specific for malignant prostate tissue.

Furthermore, database searches in public databases of the national center for biological information (NCBI) revealed the existence of several expressed sequence tags (EST clones) being in part identical to the Trp10 sequence. These EST clones were originally isolated from cancer tissues of lung, placenta, prostate and from melanoma. These clones include the clones with the following accession numbers: BE274448, BE408880, BE207083, BE791173, AI671853, BE390627. The results demonstrate that cancer cells of these tissues express Trp10 related transcripts whereas no expression of Trp10 transcripts in the corresponding healthy tissues are detectable (Figure 13). Furthermore, it could be shown that in cancer cells of melanoma and prostate cancer Trp10 transcripts are expressed as shown by in situ hybridizations using 4 antisense probes (Figure 14A – E and 13K-O and Table 2, above). Furthermore, it could clearly be shown that cancer cells of these tissues expressing Trp10 transcripts also express Trp10-antisense transcripts as shown in Figure 14F-J, Figure 14P-R and Figure 14T by in situ hybridizations using 4 sense probes (Table 2, above). The in situ hybridization experiments demonstrate that detection of a subset of cancer cells derived from carcinoma of lung, placenta, prostate and melanoma is feasible using antisense as well as sense probes complementary to Trp10 transcripts or complementary to Trp10-antisense transcripts, respectively.

The foregoing is meant to illustrate but not to limit the scope of the invention. The person skilled in the art can readily envision and produce further embodiment, based on the above teachings, without undue experimentation.

#### What Is claimed Is:

1. An isolated nucleic acid molecule encoding the human prostate carcinoma associated protein Trp8a, Trp8b, Trp9, Trp10a or Trp10b or a protein exhibiting biological properties of Trp8a, Trp8b, Trp9, Trp10a or Trp10b and being selected from the group consisting of

- (a) a nucleic acid molecule encoding a protein that comprises the amino acid sequence depicted in Figure 7, 8A, 9, 10 or 11;
- (b) a nucleic acid molecule comprising the nucleotide sequence depicted in Figure 7, 8A, 9, 10 or 11;
- (c) a nucleic acid molecule included in DSMZ Deposit No. DSM 13579, DSM 13580, DSM 13584, DSM 13581 or DSM....;
- (d) a nucleic acid molecule which hybridizes to a nucleic acid molecule specified in (a) to (c);
- (e) a nucleic acid molecule the nucleic acid sequence of which deviates from the nucleic sequences specified in (a) to (d) due to the degeneration of the genetic code; and
- (f) a nucleic acid molecule, which represents a fragment, derivative or allelic variation of a nucleic acid sequence specified in (a) to (e).
- 2. A recombinant vector containing the nucleic acid molecule of claim 1
- 3. The recombinant vector of claim 2 wherein the nucleic acid molecule is operatively linked to regulatory elements allowing transcription and synthesis of a translatable RNA in prokaryotic and/or eukaryotic host cells.
- 4. A recombinant host cell which contains the recombinant vector of claim 3.
- 5. The recombinant host cell of claim 4, which is a mammalian cell, a bacterial cell, an insect cell or a yeast cell.
- 6. An isolated protein exhibiting biological properties of the human prostate carcinoma associated protein Trp8a, Trp8b, Trp9, Trp10a or Trp10b which is encoded by a nucleic acid molecule of claim 1.
- 7. A recombinant host cell that expresses the isolated protein of claim 6.

8. A method of making an isolated protein exhibiting biological properties of the human prostate carcinoma associated protein Trp8a, Trp8b, Trp9, Trp10a or Trp10b comprising: (a) culturing the recombinant host cell of claim 6 under conditions such that said protein is expressed; and

- (b) recovering said protein.
- 9. The protein produced by the method of claim 8.
- 10. An antisense RNA sequence characterized in that it is complementary to an mRNA transcribed from a nucleic acid molecule of claim 1 or a part thereof and can selectively bind to said mRNA or part thereof, said sequence being capable of inhibiting the synthesis of the protein encoded by said nucleic acid molecule.
- 11. A ribozyme characterized in that it is complementary to an mRNA transcribed from a nucleic acid molecule of claim 1 or a part thereof and can selectively bind to and cleave said mRNA or part thereof, thus inhibiting the synthesis of the protein encoded by said nucleic acid molecule.
- 12. An inhibitor characterized in that it can suppress the activity of the protein of claim 6.
- 13. A method for diagnosing a prostate carcinoma which comprises contacting a target sample suspected to contain the protein Trp8a, Trp8b, Trp10a and/or Trp10b or the Trp8a, Trp8b, Trp10a and/or Trp10b encoding mRNA with a reagent which reacts with Trp8a, Trp8b, Trp10a and/or Trp10b or the Trp8a, Trp8b, Trp10a and/or Trp10b encoding mRNA and detecting Trp8a, Trp8b, Trp10a and/or Trp10b or the Trp8a, Trp8b, Trp10a and/or Trp10b encoding mRNA.
- 14. The method of claim 13, wherein the reagent is a nucleic acid.
- 15. The method of claim 13, wherein the reagent is an antibody.
- 16. The method of claim 13, wherein the reagent is detectably labeled.

17. The method of claim 16, wherein the label is selected from the group consisting of a radioisotope, a bioluminescent compound, a chemiluminescent compound, a fluorescent compound, a metal chelate, or an enzyme.

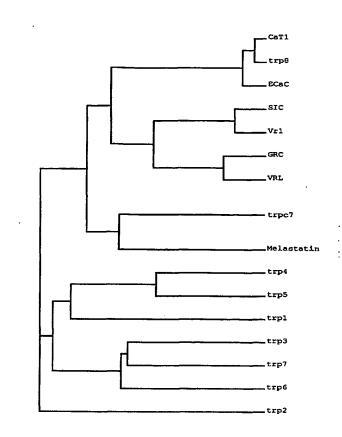
- 18. A method for diagnosing an endometrial cancer (carcinoma of the uterus) which comprises contacting a target sample suspected to contain the protein Trp8a and/or Trp8b or the Trp8a and/or Trp8b encoding mRNA with a reagent which reacts with Trp8a and/or Trp8b or the Trp8a and/or Trp8a and/or trp8b encoding mRNA and detecting Trp8a and/or Trp8b or the Trp8a and/or Trp8b encoding mRNA.
- 19. The method of claim 18, wherein the reagent is a nucleic acid.
- 20. The method of claim 18, wherein the reagent is an antibody.
- 21. The method of claim 18, wherein the reagent is detectably labeled.
- 22. The method of claim 21, wherein the label is selected from the group consisting of a radioisotope, a bioluminescent compound, a chemiluminescent compound, a fluorescent compound, a metal chelate, or an enzyme.
- 23. A method for diagnosing a melanoma, chorion carcinoma, cancer of the lung and of the prostate in a tissue of a subject, comprising contacting a sample with a reagent which detects Trp10a and/or Trp10b antisense RNA or Trp10a and/or Trp10b related antisense RNA.
- 24. A method for preventing, treating, or ameliorating a prostate tumor, endometrial cancer (carcinoma of the uterus) tumor, a chorion carcinoma, cancer of the lung or melanoma, which comprises administering to a mammalian subject a therapeutically effective amount of a reagent which decreases or inhibits expression of Trp8a, Trp8b, Trp10a and/or Trp10b and/or the activity of Trp8a, Trp8b, Trp10a and/or Trp10b.
- 25. The method of claim 24, wherein the reagent is a nucleotide sequence comprising an antisense RNA.

26. The method of claim 24, wherein the reagent is a nucleotide sequence comprising a ribozyme.

- 27. The method of claim 24, wherein the reagent is an inhibitor of Trp8a, Trp8b, Trp10a and/or Trp10b.
- 28. The method of claim 27, wherein the reagent is an anti-Trp8a-, anti Trp8b-, anti-Trp10a-and/or anti-Trp10b antibody or a fragment thereof.
- 29. A diagnostic kit useful for the detection of Trp8a, Trp8b, Trp10a and/or Trp10b or Trp8a, Trp8b, Trp10a and/or Trp10b encoding mRNA or Trp10a and/or Trp10b antisense transcripts in a sample, wherein the presence of an increased concentration of Trp8a, Trp8b, Trp9, Trp10a and/or Trp10b or Trp8a, Trp8b, Trp9, Trp10a and/or Trp10b encoding mRNA or Trp10a and/or Trp10b antisense transcripts is indicative for a prostate tumor, endometrial cancer (cancer of the uterus) tumor, a chorion carcinoma, cancer of the lung or melanoma, said kit comprising a probe for detection of Trp8a, Trp8b, Trp9, Trp10a or Trp10b or Trp8a, Trp8b, Trp9, Trp10a and/or Trp10b antisense transcripts.
- 30. The kit of claim 29, wherein the target component to be detected is Trp8a, Trp8b, Trp9, Trp10a and/or Trp10b and the probe is an antibody.
- 31. A method for identifying a compound which acts as an agonist or antagonist on the ion channels Trp8, Trp9 and/or Trp10, said method comprising contacting a test compound with the ion channel Trp8, Trp9 and/or Trp10, and determining whether said test compound affects the calcium uptake.

Figs. 1A and 1B





# В

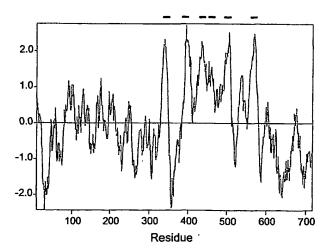
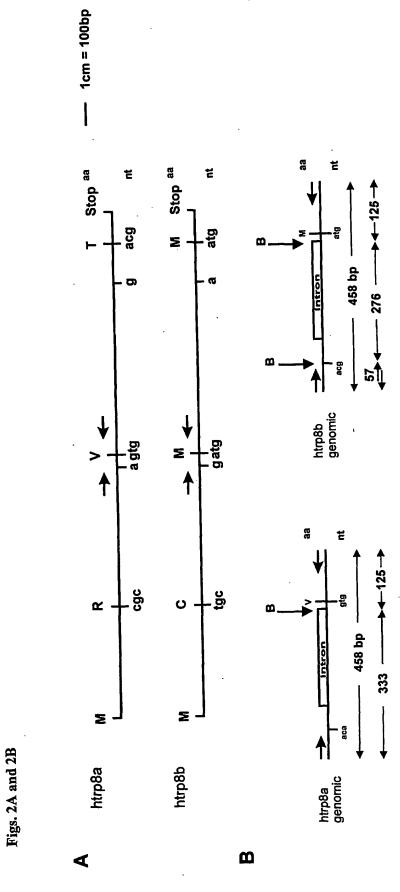


Fig. 1C

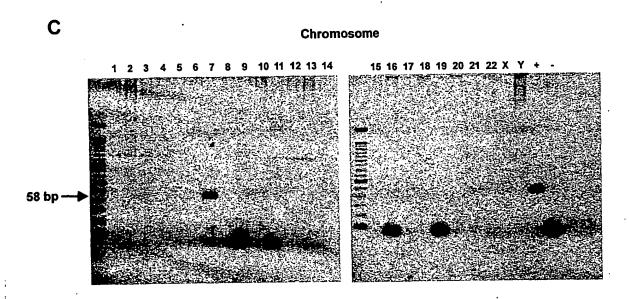
С

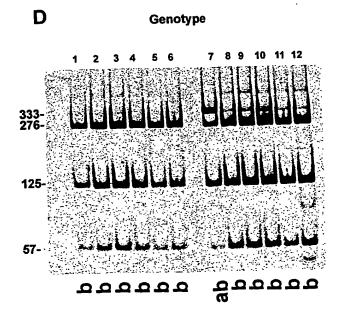
| htrp8A                       | MG  | 2          |
|------------------------------|---|------------|
| htrp&B<br>Vr1<br><i>ECaC</i> | MG<br>MEQRASLDSEESESPPQENSCLDPPDRDPNCKPPPVKPHIFTTRSRTRLFGKGDSEEASP<br>MG  | 60<br>2    |
| hat anno D.N.                | TOT THE TOTAL TH  | 52         |
| htrp8A<br>htrp8B             | LSLÜKERGLILCLMSKECKNEDIRESNADSKÜSDALLOGK-ÜLVESP-LLLÜ<br>LSLÜKERGLILCLMSKECHNEDERESNADSKÜSDALLOGK-ÜLVESP-LLLÜ<br>LOCÜ PERĞGLASCP I I YVSSVLTIGNEDGEPASVAPSOĞISVSAGEKPERLYRENSI TEN   | 52<br>52   |
| Vrl                          | LDC#YEDGLASCPIITVSSVLTIQRPGDGPASVRPSSQUSVSAGEKPPRLYDRRS1FDA   | 120        |
| ECaC                         | ACPĒKĀVĒ PRADIOKULISMPVŒ COMECYRĒRVANICOCE-ĒIRDSP-LLOĒ  | 52         |
| htrp8A                       | akdridvoaľnýcíkyedckveorganýsťašihtiáať-yds learnymeas<br>akdridvoaľnýcíkyedckveorganýsťakhtiásť-yds learnymeas   | 102        |
| htrp8B                       | AKDNOVOALNIGILKYEDCKVHORGANIETTAKHTIBAT-YDH-LEARNYIMEAR   | 102        |
| Vri<br>ECeC                  | VAOSNODERSÄLPFIORSKRUITDSEFKDERTÄNGCILKAMTALIRGONDTALLLDIVA<br>AKENDLRIÄKILLUNOSODFO ORGANGETÄNHVÄÄÄ-YTÄ LENATILMENÄ  | 180<br>102 |
| htrp8A                       |   | 156        |
| htrp8B                       | PELVTEIMTSELYEGTÄLÄTÄVVINGMOLVRALLARVÄSTÄVÄTTTÄTREP<br>RKTISLAGFVASYTISTYKIGTALÄTA ERRÄNTÄVTLÄVENGÄDIGÄVÄVÄSEFEKKTK   | 156        |
| Vr1                          | RKTDSLKOFVNASYTDSYYKGYTALHIA IERRAMTYYTLLVENGADYGLAAKGDEFKKTK   | 240        |
| ECaC                         |   | 156        |
| htrp8A                       | - MITELEGRATELY AND METERS - POLYGODZI WAS HIGHOD   | 207        |
| htrp8B<br>Vr1                | - NILITE FERRICA PAGE VINEET VILITERS - ADIRACIS LENT VILITERIO PAGE VILITERIO PA  | 207        |
| ECaC                         | MITETERIES FANONSES ON LICENS - NO RECORDS STATE IN TOP<br>MITTERS FANONSES ON LICENS - NO RECORDS FOR THE TIPO-<br>GREETE RESPECTANCE AND THE SECOND STATE OF THE TIPO-<br>HALLEY SERVICE FANCY OS SESTING THE CONTROL OF THE TIPO-<br>HALLEY SERVICE FANCY OF SESTING THE CONTROL OF THE TIPO-<br>HALLEY SERVICE FANCY OF THE TIPO-<br>THE TIPO-<br>HALLEY SERVICE FANCY OF THE TIPO-<br>THE TIPO-<br>HALLEY SERVICE FANCY OF THE TIPO-<br>HALLEY SERVICE | 300<br>207 |
| htrp8A                       |   | 261        |
| htrp8B                       | IKTEACOMMILISYDRHGUHLOPEDLYPHOGLTPFKEAGVEGYTVMFOHLYD-   | 261        |
| Vr1                          | NTKEVTSKINEILLICAKIHPTIKEELTHRKSITELALAASSEKIGVLAYILEREIHEP   | 360        |
| ECaC                         | Brièrografia in subredhioèten brozze kryere blanden de hee<br>Brièrografia i i erren en tre en i brozze kryere brande en hee<br>Brièrografia i erren en tre en i brozze kryere brande en hee<br>Brièrografia i erren en heeft brozze kryere brande en hee<br>Brièrografia i erren en heeft brande en heeft  | 261        |
| htrp8A                       | KRRHTONTYGELTSTLYDETELDSSGDEOSLIELIITTK-KREAR-QINDOTEVK   | 314        |
| htrp8B                       | KREHTOETYGELTETENDETELDSSGDEOSLEELIITTK-KREAR-QIKOOTEVK   | 314        |
| Vr1                          |   | 420        |
| ECaC                         |   | 314        |
| htrp8A                       | eīvslāmamgrpašonigairīkātāgdēmgciāmēlkpatunutsārumillogillo<br>edvslamamgrpašonigairīkātāgdīmgairkpaturutsarumilgokilo<br>rūlodnimetvariamprovēgum izšaalrēkves——Leg——-Krija<br>egvsfmakkkrappašovilaslātājāgdatociēmēlkalrodartuērditilogillo  | 374        |
| htrp8B<br>Vr1                | RILORKWIRFVKRIFYFNFFVKTIMTIFTALAYRRIVG  | 374<br>468 |
| BCaC                         | ELVSFKKKYGRPYECVLASLATISMICFTCCTPRELKLRODNRTDERDITILOOKILO  | 374        |
| •                            | Sisi  |            |
| htrp8A                       | eay troodiblygglythicallillvevpdifragytrffgotilggpfhyllityaf  | 434        |
| htrp8B                       | FAVE TPRODUCT AND TAKEN TO THE TRANSPORT OF THE PROPERTY OF TH  | 434        |
| Vr1                          | EVGUYERVTGELLSVSEGVYPFFRGTOYFLORRPSLKSLEUDSVSETLEEVOSL  | 522        |
| ECaC                         | EAYVIHODNIBLVELLVIVITAVIILLLEIPDIFRVGASRYFGGTILGGPFHVIIITTAS  92  93  | 434        |
| htrp8A                       | HVŽVIMĮNILISASCEŲVIMSĘAŽVIŽEJCNVKŲ TARGĘDILĘPFTILIŽCIMĮ FCIDINĘEC   | 494        |
| htrp8B                       | MVLVTNVMRLISASGEVVPMSEALVIGHONVMFFARGPONLGPFTTHTOROTFFGOTMRFC   | 494        |
| Vrl                          | Provintegorke varivéria aktivation de l'avaite de l'avaite de l'avaite de l'avaite de l'avaite de l'avaite de l<br>L'alliment invage vollé versité de l'avaite de l'avaite de l'avaite de l'avaite de l'avaite de l'avaite de l'a   | 582        |
| <b>ECaC</b>                  | TATELLINGUETUM MEDIA PER PARTE  | 494        |
| htrp8A                       | WINATVILIGEASAFYIIFOTED   | 538        |
| htrp8B                       | winaviilēasāfyiifoted——eeelō-hfydypmalfstāelv<br>winaviilēasāfyiifoted——eeelō-hfydypmalfstāelv  | 538        |
| Vrl                          | fvylvilfligistävviliedgknislemestphkcrgsackpensynslystclelikft  | 642        |
| ECaC                         | wimayvilgeasafhitfoted  | 538        |
| 1                            | a marking property and a state of the state   |            |
| htrp8A<br>htrp8B             | LTITICOANYNYDLPFMYSITYAAFAI LATIJMIRLIJAAMIGIEHNKVAHERIEIMRADI  | 598<br>598 |
| Vr1                          | IGGELEFTENYDFKAVFITLLLAYVELTYTILLANGERI MEETVIKTROESKNING OR  | 702        |
| <b>DCaC</b>                  | LTIIŪGPANYAVŪLPPHYS ĮTYKĀFAĮ LATLĀTĀJLIĀMĀGDĒHHKVĀLĒKDELĀRAĞI<br>LTIIŪGPANYAVŪLPPHYS ĮTYKĀFAJ LATLĀTĀJLIĀMĀGDĪHKRVĀLĒKDELĀRAĞI<br>IGNGOLEPTSNIPFKAVPIQLIĀNYĀLYTŪLĀNAĀ LUSEVANKIĀGSKAULĀG<br>LTIIŪGPANYSVĀLPPHYC ĮTYAĀFAJ LATLĀCĀĻLFĒMĀGIDHHKVĀGĒRUDEJĀRAŅV  | 598        |
|                              | S6  |            |
| htrp8A                       | VATTVOLERGLERCLME—RSS——ICHESYGLGD—RHILEGEBRODLARGRIGHVADA<br>VAXTVOLERGLERCLME—RSS——ICHESYGLGD—RHILEGEBRODLARGRIGHVADA<br>ALGILADERSFLACHRAFEGGLLOVETTEORGDYRHCERYDEVMITTHRITHVGI INE   | 671        |
| htrp8B<br>Vrl                | ATTILITERSFLETMERAFESCRITOREFFECENTIAL CONTROL OF THE CONTROL OF T  | 671<br>762 |
| ECaC                         | VASTVALERKAPRELAP—ESS——ICEYEYGLGD—KAFILKYENHHDONPLRVLRYVEA  | 671        |
| htrp8A                       | FHTR  | 726        |
| htrp8B                       | FHTRESEDLDKDSV-EKLELGCPFSPHLSLP-PSVSRSTSRSSANMERLROGTLRR  | 726        |
| Vr1<br>ECaC                  | DPGN-CEGVKRTLSFSLRSGRVSGRNNKNFALVPLLRDASTRURHATOGEEVOLKHYTG   | 820        |
|                              | FKCSDKEDGQEQLSEKRP-STVESGMLSRASVAFQTPSLSRTTSQSSN—SHRGWEILRR   | 728        |
| htrp8A<br>htrp8B             | Dlagi inrgledgesweyqi* Dlagi inrgledgesweyqi*   | 746<br>746 |
| Vrl                          | SLKPEDAEVFKDSHVPGEK*  | 839        |
| ECaC .                       | ntighinigidigegdgeevyhp•  | 751        |



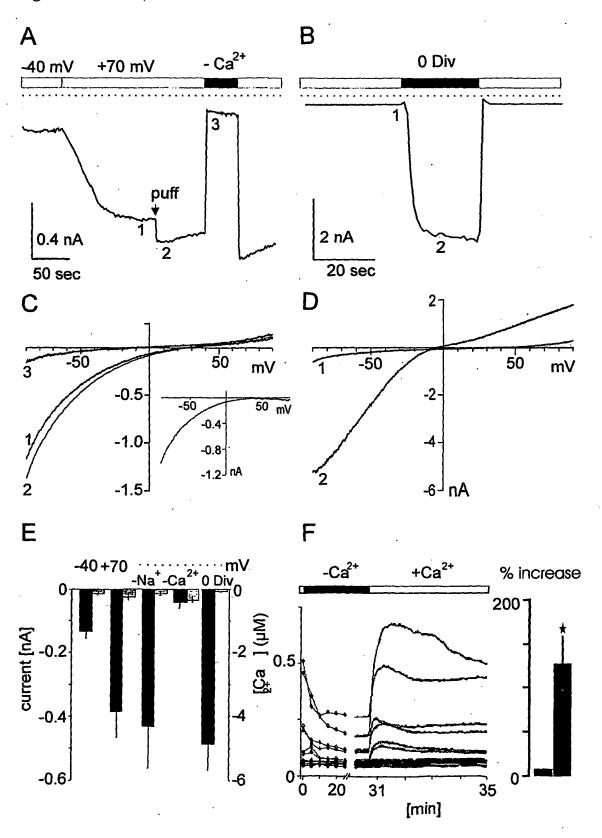
**SUBSTITUTE SHEET (RULE 26)** 

Figs. 2C and 2D

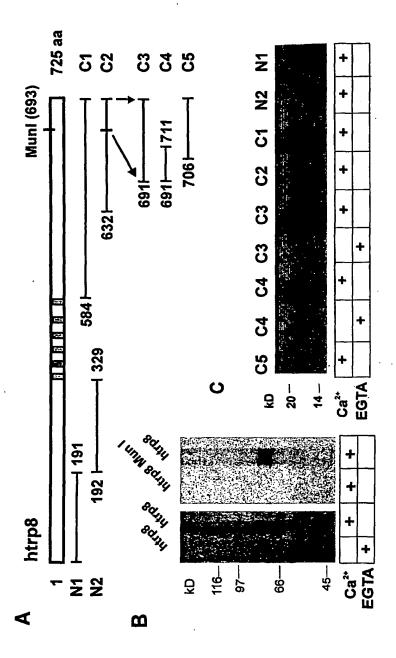




Figs. 3A - 3F



SUBSTITUTE SHEET (RULE 26)



Figs. 4A - 4C

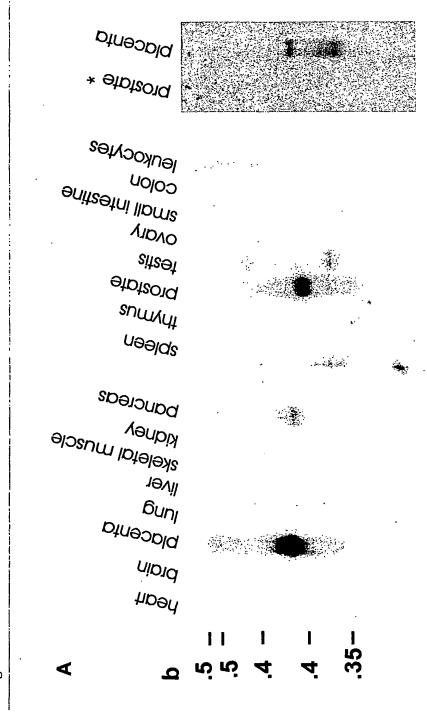
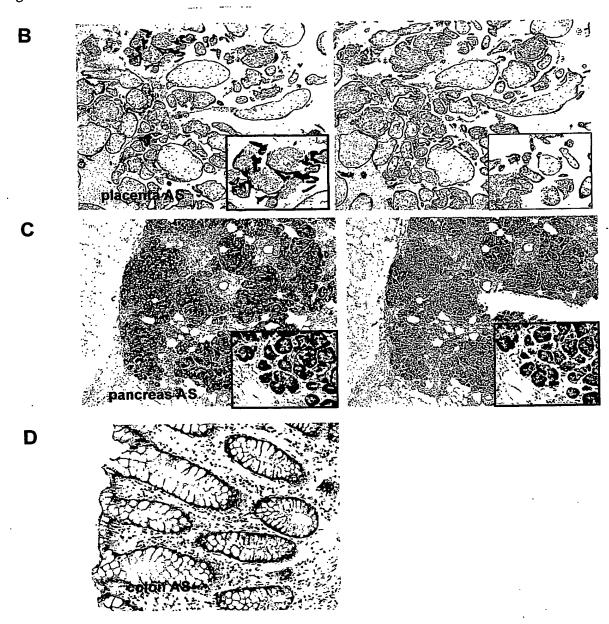


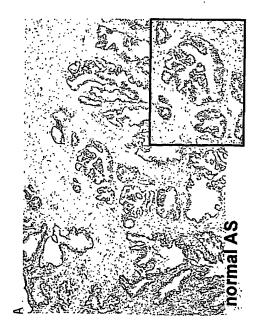
Fig. 5A

Figs. 5B - 5D

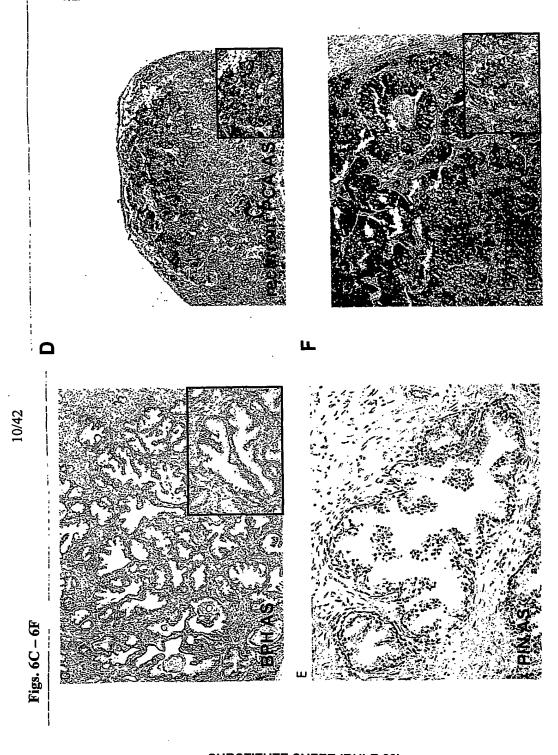


Primary PCA

 $\mathbf{\omega}$ 



Figs. 6A and 6B



**SUBSTITUTE SHEET (RULE 26)** 

PCT/EP01/08309

|                |             |        | 10            |       |           |              |       |        | 30        |           |            |           |             |       | 5           | 0         |   |                  |      |
|----------------|-------------|--------|---------------|-------|-----------|--------------|-------|--------|-----------|-----------|------------|-----------|-------------|-------|-------------|-----------|---|------------------|------|
| -              | GCCZ        | AAG:   | rgtaac        | CAAAC | CTC       | ACA(         | SCC   | CTC    | TCCA      | AA        | CTG        | CTC       | GGG         | CTC   | CTG         | GGA       | GAC   | TC               | CCA  |
|                |             |        | 70            |       |           |              |       |        | 90        |           |            |           |             |       | 11          | 0         |   |                  |      |
|                | AGG!        | AAC'   | rcgtc?        | AGGA  | AGG       | CAG          | GAG   | ACA    | GGAG      | ACC       | GGG        | ACC'      | CT          | ACA   | GGA         | GAC       | GG1   | 'GG(             | SCC  |
| •              |             |        | 130           |       |           |              |       |        | 150       |           |            |           |             |       | 17          | _         |   |                  |      |
|                | GGC         | CCT'   | TGGGGG        | GGC   | TGA'      | rgt(         | GGC   | CCC    | AAGC      | CT        | GAG:       | rcc       | CGT         | CAG   |             |           | CCI   | CG               | 3CC  |
|                |             |        | 190           |       |           |              |       |        | 210       |           |            |           |             |       | 23          |           |   |                  |      |
|                | TCA         | GGC    | CCCCA         | AGGA  | GCC       | GGC          | CCT   | ACA    | ccc       | CAT       | GGG:       | rtt       |             |       |             |           |   |                  |      |
|                |             |        |               |       |           |              |       |        |           | M         | G          | L         | S           | L     | P           |           | E   | K                | G    |
| Fig. 7         |             |        | 250           |       |           |              |       |        | 270       |           |            |           |             |       | 29          | -         |   |                  |      |
| . <u>rg.</u> / | GCT         | TAA    | TCTCT         |       |           |              |       |        |           |           |            |           |             |       |             |           |   |                  |      |
|                | L           | Ι      | r c           | L     | W         | S            | K     | F      | С         | R         | W          | F         | Q           | R     | R           |           | S   | W                | A    |
|                |             |        | 310           |       |           |              |       |        | 330       | ~~~       | ~~~        | C7.C      | ~ 3 M       | ome   | 35<br>CCD C |           | ייטריי  | Tr <b>CT</b> Tr  | CCTI |
|                |             |        | CCGAG         |       |           |              |       |        | Q<br>Q    |           |            |           |             |       |             | S         | P   | L                |      |
|                | Q           | S      | R D<br>370    | E     | Q         | N            | ь     |        | 390       | Q         | Λ.         | K         |             | **    |             | LO        | -   |                  | _    |
|                | un/cut      | አርር    | TGCCA         | משמת  | ጥልል       | ·<br>TCD     | ጥርጥ   |        |           | CCT       | CAA        | CAA       | ርሞጥ         | сст   |             |           | rga   | GGA              | TTG  |
|                |             |        | A K           |       |           |              |       |        | A         |           |            |           |             |       |             |           | E   | D                |      |
|                | -           | •      | 430           |       | ••        | _            | •     | ~      | 450       | _         | •••        |           | _           |       |             | 70        |   |                  |      |
|                | CAA         | GGT    | GCACC         | AGAG  | AGG       | AGC          | CAT   | GGG    | GGA       | AAC       | AGC        | GCT       | ACA         | CAT   | AGC         | AGC       | CCT   | CTA              | TGA  |
|                | K           |        |               |       | G         |              | M     |        | E         |           |            |           | н           | I     |             |           |   | Y                | D    |
|                |             |        | 490           |       |           |              |       |        | 510       |           |            |           |             |       | 53          | 30        |   |                  |      |
|                | CAA         | CCT    | GGAGG         | CCGC  | CAT       | 'GGT         | GCT   | 'GAI   | rgga      | GGC       | TGC        | CCC       | GGA         | GCT   | 'GGT        | CTT       | TGA   | GCC              | CAT  |
|                | N           | L      | E A           | . A   | M         | V            | L     | M      | E         |           | A          | P         | E           | L     |             | _         | E   | P                | M    |
|                |             |        | 550           |       |           |              |       |        | 570       |           |            |           |             |       | -           | 90        |   |                  |      |
|                |             |        | TGAGC         |       |           |              |       |        |           |           |            |           |             |       |             |           |   |                  |      |
|                | T           | S      | E L           | Y     | E         | G            | Q     | T      | A         |           | Н          | I         | A           | V     | ۷           | N<br>50   | Q   | Ð                | M    |
|                | <b>67.7</b> | ~~     | 610<br>GGTGC  | an ad | 100       |              | m     |        | 630       |           | יראר       | an Can    | C-mC        | anc.c | _           |           | ראר   | ישכני            | CAC  |
|                | GAA         |        | V R           |       | L         |              |       |        | R<br>R    |           |            |           |             |       |             |           | T   | дос<br>G         | T    |
|                | N           | ъ      | 670           | . A   | n         | 1            | ^     |        | 690       |           |            | •         | ٥           | ••    |             | 10        | -   | •                | _    |
|                | TGC         | CTT    | CCGCC         | GTAG  | TCC       | :CCG         | CAA   | CC     |           |           | CTI        | TGG       | GGA         | GCA   | CCC         | TTT       | GTC   | CTT:             | TGC  |
|                |             | F      |               |       | P         | R            | N     | L      |           | Y         |            |           |             |       | P           |           | S   |                  | A    |
|                |             |        | 730           |       |           |              |       |        | 750       |           |            |           |             |       |             | 70        |   |                  |      |
|                | TGC         | CTG    | TGTGA         | ACAC  | TGP       | LGG#         | GAT   | 'CG    | rgcg      | GCI       | GCI        | CAT       | TGA         | GCF   | TGG.        | AGC       | TGA   |                  |      |
|                | A           | С      | V N           | S     | E         | E            | I     | V      | R         |           | Г          | I         | E           | Н     | _           | A         | D   | I                | R    |
|                |             |        | 790           |       |           |              |       |        | 810       |           |            |           |             | ~     | _           | 30<br>222 |   |                  |      |
|                |             |        | LGGACT        |       |           |              |       |        |           |           |            |           |             |       |             |           |   | K                |      |
|                | A           | Q      |               | L     | G         | N            | T     | ٧      | L<br>870  |           | Ţ          | L         | I           | r     | Ō           | 90        | N   | Α.               | •    |
|                | ادرات       | ነጥር ር  | 850<br>CTGCC  | ימכמי | рсти      | יראז         | ירנים | مانتاء |           |           | נייטי      | CG        | CAC         | AC    | _           | _         | CCZ   | CCI              | rgca |
|                |             |        | CCC           |       | Y         |              |       |        | L         | S         |            | D         | R           | H     |             | D         | н   | L                | Q    |
|                | -           | •      | 910           | 2 13  | •         | -            | ~     |        | 930       | _         | -          | -         |             |       |             | 50        |   |                  | _    |
|                | GCC         | CCT    | GGACC         | TCG   | rgco      | CA           | ATC   | (CC    |           |           | CAC        | ccc       | TT          | CAZ   | AGCT        | GGC       | TGG   | AG               | rgga |
|                | P           | L      | D I           | v     | P         | N            | Н     | Q      | G         | L         | T          | P         | F           | K     | L           | A         | G   | V.               | E    |
|                |             |        | 970           |       |           |              |       |        | 990       |           |            |           | -           |       |             | 10        |   |                  |      |
|                |             |        | CACTO         |       |           |              |       |        |           |           |            |           |             |       |             |           |   |                  |      |
|                | G           | N      | T V           | M     | F         | Q            | H     | L      | M         | Q         | K          | R         | K           | H     | T           | Q         | W   | T                | Y    |
|                |             |        | 1030          |       |           |              |       |        | 1050      |           |            |           |             |       |             | 70        |   |                  |      |
|                |             |        | ACTGA         |       |           |              |       |        |           |           |            |           |             |       |             |           |   |                  |      |
|                | G           | Þ      | L T           |       | T         | L            | Y     |        |           |           | E          | I         | D           | S     |             |           | D   | E                | Q    |
|                |             |        | 1090          |       |           |              |       |        | 1110      |           |            |           |             |       |             | .30       |   |                  |      |
|                |             |        | CCTGG         |       |           |              |       |        |           |           |            |           |             |       |             |           |   |                  |      |
|                | S           | L      | L E           |       | 1         | 1            | T     |        |           |           | R          | E         | A           | K     |             | 90        | r   | U                | v    |
|                |             |        | 1150          |       |           | N-C-C-       |       |        | 1170      |           | ~~» 1      |           | ·cm         |       |             |           | יים ארבויי<br>איים איים איים איים איים איים איים אי | 7 Cdr.           | rCTG |
|                | GAC.        | אטע    | CGGTGA<br>V K | MGG   | 36C)      | [יטטו<br>••• | GA(   | JUU'   | TCAA      | UTU<br>ta | 140c<br>17 | אטטג<br>מ | V<br>V      | ころら   | 9906<br>D   | D         | V<br>V  | r<br>V           | . C  |
|                | Т           | Ľ      |               | E     | ь         | ٧            | S     |        |           |           | K          | K         | 1           | G     |             | 50        | 1   | Ľ                | _    |
|                | (13.00      | ما عاد | 1210<br>GGGTG | יעראי | יוחקין    | ነጥ ሶሳ        | ייכיי |        | 1230      |           | րլու       | اسات      | ימיי        | ימיטר |             |           | ימי   | rcm <sup>z</sup> | ACCG |
|                | CAT<br>M    | T      | G A           | T     | A<br>TUTA | 7 L.C.       | . GC] | UL.    | nuni<br>T | ⊬M1<br>T  | ب<br>م     | L<br>L    | φ.<br>- (C) | M     | . O. C      | <br>C     | T   | Υ                | R    |
|                | PI          | ם      | 1270          |       | 1         | 11           | ъ     |        | 1<br>1290 |           | C          | Ľ         | ,           |       |             | 10        | -   | -                |      |
|                |             |        | 1210          |       |           |              |       |        | 1230      | •         |            |           |             |       |             |           |   |                  |      |

12/40

| CCC | CCT      | CAAG        | CCC  | AGG  | ACC  | TAA:      | 'AA'     | CCGC | CACA       | AGC       | ccc   | CGG      | GAC          | AAC       | ACC       | CTC       | TTP        | CAG       | CA        |
|-----|----------|-------------|------|------|------|-----------|----------|------|------------|-----------|-------|----------|--------------|-----------|-----------|-----------|------------|-----------|-----------|
| P   |          | ĸ           |      |      |      |           |          |      | T          |           | P     | R        | D.           | N         | T         | L         | L          |           |           |
|     |          | 1330        |      |      |      |           |          |      | 350        |           |       |          |              |           | 137       | -         |            |           |           |
| GAA | GCT      | ACTI        | CAG  | GAA  | GCC  | TAC       | CGT      | SAC  | CCCI       | 'AAC      | GAC   | GA?      | TAT          | CGG       | CIC       | GT(       | CGG        | GAG       | CT        |
| K   |          | L           |      |      |      |           |          |      | P          |           |       |          |              |           | L         |           |            | E         |           |
|     |          | 1390        | )    |      |      |           |          | 1    | 410        |           |       |          |              |           | 143       | 0         |            |           |           |
| GGT | GAC      | TGTC        | ATT  | rGGG | GC:  | rat(      | CAT      | CAT  | CCTC       | CTO       | GGT   | AGA      | GGT"         | rccz      | AGA(      | TAC       | CTTC       | CAG       | TA        |
| V   |          | V           |      |      |      |           |          |      | L          |           |       |          |              |           |           |           | F          | R         | M         |
|     |          | 1450        | )    |      |      |           |          | 1    | 470        |           |       |          |              |           | 149       | 90        |            |           |           |
| GGG | GGI      | CACI        | CGC  | CTT  | TT   | rggi      | ACA      | GAC  | CATO       | CT        | TGG   | GGG      | cca          | ATT       | CA:       | rgt       | CCT        | CATO      | TAC       |
|     |          | T           |      |      |      |           |          |      | I          |           |       |          |              |           |           |           | L          |           | 1         |
|     | -        | 1510        | )    | _    |      |           | _        |      | 530        |           |       |          |              |           | 15        |           |            |           |           |
| CAC | CTA      | TGC         | TTC  | CATO | GT(  | GCT       | GGT      | GAC  | CAT        | GGT       | GAT   | GCG      | GCT          | CAT       | CAG'      | rgc       | CAG        | CGGC      | <b>GA</b> |
| T   |          | A           |      |      |      |           |          |      | M          |           |       |          |              |           |           |           | S          | G         | E         |
|     |          | 1570        | כ    |      |      |           |          |      | 590        |           |       |          |              |           | 16        | 10        |            |           |           |
| GGT | GGT      | ACC         | CAT( | GTO  | TT   | TGC       | ACT      | CGT  | GCT        | GGG       | CTG   | GTG      | CAA          | CGT       | CAT       | GTA       | CTT        | CGC       | CCG       |
|     |          | P           |      |      |      |           |          |      | L          |           |       |          |              |           |           |           |            | A         |           |
|     |          | 1630        | 0    |      |      |           |          |      | 650        |           |       |          |              |           | 16        |           |            |           |           |
| AGG | :ATT     | CCA         | GAT  | GCT  | AGG  | ccc       | CTT      | CAC  | CAT        | CAT       | GAT   | TCA      | GAA          | GAT       | GAT'      | TTT       | TGG        | CGA       | CCT       |
| G   |          | Q           |      |      |      |           |          |      |            |           |       |          |              |           |           |           | G          | D         |           |
|     |          | 169         |      |      |      |           |          |      | 710        |           |       |          |              |           | 17        | 30        |            |           |           |
| GAT | GCC      | ATT         | ctg  | CTG  | GCT  | GAT       | GGC      | TGT  | GGT        | CAT       | CCT   | GGG      | CTT          | TGC       | TTC       | AGC       | CTT        | CTA       | TAT       |
| М   |          | F           |      | M    |      |           |          |      | v          |           |       |          |              |           |           | A         | F          |           |           |
|     |          | 175         | 0    |      |      |           |          | 1    | 770        |           |       |          |              |           | 17        | 90        |            |           |           |
| CAT | CT       | rcca        | GAC  | AGA  | GGA  | ccc       | CGF      | GGA  | GCT        | AGG       | CCA   | CTI      | CTA          | CGA       | CTA       | CCC       | CAT        | GGC       | CCT       |
| I   |          | Q           |      |      |      |           |          |      | L          |           |       |          |              |           |           |           | M          |           | L.        |
|     |          | 181         |      |      |      |           |          |      | 830        |           |       |          |              |           |           | 50        |            |           |           |
| GTT | CAC      | GCAC        | CTT  | CGA  | GCT  | GTT       | CCI      | TAC  | CAT        | CAT       | CGP   | TGG      | CCC          | AGC       | CAA       | CTA       | CAA        | CGT       | GGA       |
| F   | S        | T           | F    | E    | L    | F         | L        | T    | I          | I         | D     | G        | P            | A         | N         | Y         | N          | V         | D         |
|     |          | 187         |      |      |      |           |          |      | 890        |           |       |          |              |           |           | 10        |            |           |           |
| CC  | rgc      | CCTT        | CAT  | GTA  | CAG  | CAT       | CAC      | CTF  | TGC        | TGC       | CTI   | TGC      | CAT          | CAT       | CGC       | CAC       | ACT        | GCT       | CAT       |
| L   | P        | F           | M    | ¥    | S    | I         | Ŧ        | Y    | A          | A         | F     | A        | I            | I         |           | T         | L          | L         | M         |
|     |          | 193         |      |      |      |           |          |      | .950       |           |       |          |              | •         |           | 70        |            |           |           |
| GC: | rca.     | ACCT        | CCT  | CAT  | TGC  |           |          |      |            |           |       |          |              |           |           |           |            |           |           |
| r   | N        | L           | L    | 1    | A    | M         | M        |      | Đ          |           | H     | W        | R            | V         |           |           | Ε          | R         | D         |
|     |          | 199         |      |      |      |           |          |      | 2010       |           |       |          |              |           |           | 30        |            |           |           |
|     |          | TGTG        |      |      |      |           |          |      |            |           |       |          |              |           |           |           |            |           |           |
| E   | L        | W           |      | A    | Q    | I         | V        |      | T          |           | V     | M        | L            | B         | R         |           | ь          | P         | K         |
|     |          | 205         |      |      |      |           |          |      | 2070       |           |       |          |              |           |           | 90        | . ~~       |           | - COM     |
|     |          | TGTG        |      |      |      |           |          |      |            |           |       |          |              |           |           |           |            | W         | F         |
| C   | L        | W           |      | R    | S    | G         | I        |      | G          |           | B     | Y        | G            | ъ         |           |           | K          | w         | £         |
|     |          | 211         |      |      |      |           |          |      | 2130       |           |       |          |              |           |           | 50        | . ~~       | · N C 7   | CCC       |
|     |          | GGGT        |      |      |      |           |          |      |            |           |       |          |              |           |           | Y         |            |           |           |
| Ļ   | R        | V           |      | D    | R    | Q         | ע        |      | N          |           | Ų     | K        | 1            | Q         |           | 210       | n          | ¥         | n         |
|     |          | 217<br>ACAC |      |      | -cm- | <b></b> . |          |      | 2190       |           | 1 7 C | N CADA   | 77 C         | ncci      |           |           | ים<br>דמרי | cca       | יבכב      |
|     | rcc      | ACAC<br>T   | CCG  | GGG  | CTU  | AG/       | iGG:     | ATT. | ruu.       | ichi<br>V | THOU  | ACT      | . تاهم<br>17 | ruu.<br>F | vvvv<br>v | t.        | r<br>F     | L         |           |
| F   | н        |             |      | G    | S    | E         | ט        |      | ں<br>2250ء |           | υ     | 3        | . v          | E         |           | 270       | 15         |           | ٠         |
|     |          | 223<br>CCTT |      |      |      |           | h carrie |      |            | -         | ~~~   |          | ~ > _ <      | rcm       |           |           | מייב       | ጉርጥር      | rece      |
|     | GTC<br>- | CCTI<br>F   | CAG  | CCC  | :CC; | ACC:      | rGT      |      | rrcu       | JAI.      | المال | JU11     | - DEN        | 191       | אטני      | o<br>Para | u,<br>atte | ~~.       | 2000      |
| С   | P        |             |      | Р    | н    | т         | S        |      |            |           | P     | 5        | ٧            | 3         |           | 330       |            |           | 20        |
|     |          | 229<br>GTGC |      |      |      |           |          |      | 2310       |           | ~~ 1  | 200      | men.         | ~~~       |           |           | rcc        | -TY-2/    | таэс      |
|     |          | GTGC<br>A   |      |      |      |           |          |      |            |           |       |          |              |           |           |           |            |           |           |
| S   | S        |             |      | W    | E    | R         | ь        |      |            |           | T     | יו       | K            | K         |           | 390       |            | G         | •         |
|     |          | 235         |      |      |      |           |          |      | 2370       |           |       |          | 3 m/l        | 2021      |           |           |            | anca      | TCT       |
|     |          |             |      |      |      |           |          |      |            |           |       |          |              |           | LCI       | oac.      | 190        | 310       |           |
| I   | N        | R           |      | ъ    | E    | Ð         | G        |      |            |           | E     | 1        | Q            | 1         | •         |           |            |           |           |
|     |          | 241         |      |      |      |           |          |      | 2430       |           | -     |          | ~~-          |           | _         | 150       | 7 C 7 '    | . n. n. 1 | ACC A     |
| CA  | CTT      |             |      | TGC  | AA.  | TT(       | -CT      |      |            |           | C1'G  | GGT'     | GCA.         | TCA       |           |           |            | mm        | ACCA      |
|     |          | 247         |      |      |      |           |          |      | 2490       |           |       | <b>-</b> |              |           |           | 510       |            |           | ~~~ ^     |
| AA  | CAC      |             |      | TC   | CA.  | CT        | CCC      |      |            |           | GGA   | JAA      | AGA          | GAا       |           |           |            | H-(-)     | CCAA      |
|     |          | 253         |      | _    |      |           |          |      | 2550       |           |       |          |              |           |           | 570       |            | ~~~       | ~2~       |
| GG. | AAT      | GTAC        | GT1  | 'GAC | AA.  | CA(       | CTG      | CTC  | CAG        | 3CC       | rGC.  | ATT.     | ACT          | CCT       | rCA(      | JCT       | CIG        | أحاحات    | CAGA      |

13/40

Fig. 7 / continuation 2

| 2590                 | 2610                | 2630                   |   |
|----------------------|---------------------|------------------------|---|
| GGAAGCCCAGCCCAAGCAC  | GGGGCTGGCAGGGCGTGAG | GAACTCTCCTGTGGCCTGCTC  | A |
| 2650                 | 2670                | 2690                   |   |
| TCACCCTTCCGACAGGAGC  | ACTGCATGTCAGAGCACTT | TAAAAACAGGCCAGCCTGCTT( | 3 |
| 2710                 | 2730                | 2750                   |   |
| GGCCCTCGGTCTCCACCCC  | AGGGTCATAAGTGGGGAGA | GAGCCCTTCCCAGGGCACCCA  | 3 |
| 2770 .               | 2790                | 2810                   |   |
| GCAGGTGCAGGGAAGTGCA  | GAGCTTGTGGAAAGCGTG1 | 'GAGTGAGGGAGACAGGAACGG | C |
| 2830                 | 2850                | 2870                   |   |
| TCTGGGGGTGGGAAGTGG   | GCTAGGTCTTGCCAACTCC | ATCTTCAATAAAGTCGTTTTC  | G |
| 2890                 | 2910                |                        |   |
| GATCCCTAAAAAAAAAAAAA | АААААААААААА        |                        |   |

MGLSLPKEKGLILCLWSKFCRWFQRRESWAQSRDEQNLLQQKRIWESPLLLAAKDNDVQALNKLLKYEDCKVHQRGAMGETALHIA ALYDNLEAAMVLMEAAPELVFEPMTSELYEGQTALHIAVVNQNMNLVRALLARRASVSARATGTAFRRSPRNLIYFGEHPLSFAAC VNSEEIVRLLIEHGADIRAQDSLGNTVLHILILQPNKTFACQMYNLLLSYDRHGDHLQPLDLVPNHQGLTPFKLAGVEGNTVMFQH LMQKRKHTQWTYGPLTSTLYDLTEIDSSGDEQSLLELIITTKKREARQILDQTPVKELVSLKWKRYGRPYFCMLGAIYLLYIICFT MCCIYRPLKPRTNNRTSPRDNTLLQQKLLQEAYVTPKDDIRLVGELVTVIGAIIILLVEVPDIFRMGVTRFFGQTILGGPFHVLII TYAFMVLVTMVMRLISASGEVVPMSFALVLGWCNVMYFARGFQMLGPFTIMIQKMIFGDLMRFCWLMAVVILGFASAFYIIFQTED PEELGHFYDYPMALFSTFELFLTIIDGPANYNVDLPFMYSITYAAFAIIATLLMLNLLIAMMGDTHWRVAHERDELWRAQIVATTV MLERKLPRCLWPRSGICGREYGLGDRWFLRVEDRQDLNRQRIQRYAQAFHTRGSEDLDKDSVEKLELGCPFSPHLSLPTFSVSRST SRSSANWERLRQGTLRRDLRGIINRGLEDGESWEYQI

Figure 8:

ATGGGTTTGTCACTGCCCAAGGAGAAAGGGCTAATTCTCT MGLSLPKEKGLILC 250 270 290 GCCTATGGAGCAAGTTCTGCAGATGGTTCCAGAGACGGGAGTCCTGGGCCCAGAGCCGAG L W S K F C R W F Q R R E S W A Q S R D 350 330 ATGAGCAGAACCTGCTGCAGCAGAAGAGGATCTGGGAGTCTCCTCTCTTCTAGCTGCCA EQNLLQQKRIWESPLLLAAK 390 410 AAGATAATGATGTCCAGGCCCTGAACAAGTTGCTCAAGTATGAGGATTGCAAGGTGCACC D N D V Q A L N K L L K Y E D C K V H Q 470 450 430 AGAGAGGAGCCATGGGGGAAACAGCGCTACACATAGCAGCCCTCTATGACAACCTGGAGG R G A M G E T A L H I A A L Y D N L E A 510 530 CCGCCATGGTGCTGATGGAGGCTGCCCCGGAGCTGGTCTTTGAGCCCATGACATCTGAGC AMVLMEAAPELVFEPMTSEL 570 590 TCTATGAGGGTCAGACTGCACTGCACATCGCTGTTGTGAACCAGAACATGAACCTGGTGC Y E G Q T A L H I A V V N Q N M N L V R 630 650 GAGCCCTGCTTGCCCGCAGGGCCAGTGTCTCTGCCAGAGCCACAGGCACTGCCTTCCGCC A L L A R R A S V S A R A T G T A F R R 710 690 GTAGTCCCTGCAACCTCATCTACTTTGGGGAGCACCCTTTGTCCTTTGCTGCCTGTGTGA S P C N L I Y F G E H P L S F A A C V N

Fig. 8 / continua n 1

|  |   | 30   |   |  |   |   |   | 750  |  |   |   |  | <b></b>   | 770  |  | <i></i>  | a                                    |
|--|---|--|---|--|---|---|---|--|--|---|---|--|---|--|--|--|--------------------------------------|
|  |   |  |   |  |   |   |   | CATTGA<br>I E  | GCA<br>H   |   | agc<br>a                                  |  | CAT   | R A  |  |  |                                      |
| S  |   | E<br>90  | Ι   | V  | R   | Ь.  | ь   | 810  | н  | G   | A   | υ  | 1   | 830  | Q  | D  | S                                    |
| CCCT   | GGGA  | AAC  | CAC   | AGT                                      | GTT   | ACA   | CAT   | CCTCAT   | CCT  | CCA   | GCC                                       | CAA  | CAA   | AACCT  | PTGC   | CTG  | CC                                   |
| ь  | G<br>R  | N<br>50  | T   | ٧  | L   | Ħ   | I   | L I<br>870   | L  | Q   | P   | И  | K   | T F<br>890   | A  | С  | Q                                    |
| AGAT   |   |  | ст  | GTT                                      | GCTV  | GTC   | СТА   | CGACAG   | ACA  | TGG   | GGA                                       | CCA  | ССТ   |  | CCT  | GGA  | cc                                   |
| М  | Y   |  | _   | L  |   |   |   | D R  |  |   | D   |  | L   | Q P  | _  | D  | L                                    |
|  | 9   | 10   |   |  |   |   |   | 930  |  |   |   |  |   | 950  |  |  |                                      |
| TCGT   | GCCC  | 'AA'   | rca(  | CCA                                      | GGG   | TCT   | CAC   | CCCTTT   | CAA  | GCT   | GGC                                       | TGG  | AGT   | GGAGG  | STAA   | CAC  | TG                                   |
| V  | P   | N  | H   | Q  | G   | L   | T   | P F  | ĸ  | L   | A   | G  | v   | E G  | N  | T  | v                                    |
|  | 9   | 70   |   | _  |   |   |   | 990  |  |   |   |  |   | 1010   |  |  |                                      |
| TGAI   | GTTI  | 'CAC   | GCA(  | CCT                                      | GAT   | GCA   | GAA   | GCGGAA   | GCA  | CAC   | CCA                                       | GTG  | GAC   | GTATG  | GACC   | ACT  | GA                                   |
| M  | F   | Q  | H   | L  | M   | Q   | K   | R K  | Н  | T   | Q   | W  | T   | Y G  | P  | L  | T                                    |
|  | 10  | 30   |   |  |   |   |   | 1050   |  |   |   |  |   | 1070   |  |  |                                      |
| CCTC   | GACT  | CTC  | CTA'  | TGA                                      | CCT   | CAC   | AGA   | GÅTCGA   | CTC  | CTC   | AGG                                       | GGA  | TGA   | GCAGT  | CCCI   | 'GCT   | GG                                   |
| S  | T   | L  | Y   | D  | L   | T   | E   | I D  | S  | S   | G   | D  | E   | Q S  | L  | L  | E                                    |
|  | 10  | 90   |   |  | •   |   |   | 1110   |  |   |   |  |   | 1130   |  | ٠.   |                                      |
| AACT   | TATC  | AT(  | CAC   | CAC                                      | CAA   | GAA   | GCG   | GGAGGC   | TOC  | CCA   | GAT                                       | CCT  | GG?   | CCAGA  | CGCC   | GGT  | GΑ                                   |
| L  | I   | I  | T   | T  | K   | K   | R   | E A  | R  | Q   | Ī   | L  | D   | Q T  | P  | V  | K                                    |
|  | 11  | .50  |   |  |   |   |   | 1170   |  |   | ÷   |  |   | 1190   |  |  |                                      |
| AGGA   | GCTG  | GT   | SAG   | CCT                                      | CAA   | GTG   | GAA   | GCGGTA   | .CGG   | GCG   | GCC                                       | GTA  | CTI   | CTGCA  | TGCI   | GGG  | TG                                   |
| E  | L   | V  | S   | L  | K   | M   | ĸ   | R Y  | G  | R   | P   | Y  | F   | СМ   | L  | G  | A                                    |
|  |   | 10   |   |  |   |   |   | 1230   |  |   |   |  |   | 1250   |  |  |                                      |
|  |   |  |   |  |   |   |   | CTTCAC   |  |   |   |  |   |  |  |  | -                                    |
| I  | Y   | L  | L   | Y  | I   | I   | С   | F T  | M  | С   | С   | I  | Y   | R P  | L  | K  | P                                    |
|  |   | ?70  |   |  |   |   |   | 1290   |  |   |   |  |   | 1310   |  |  |                                      |
|  |   |  |   |  |   |   |   | CCGGGA   |  |   |   |  |   |  |  |  |                                      |
| R  | T   |  | N   | R  | T   | S   | P   | R D  | И  | T   | L   | L  | Q   | Q K  | L  | Ļ  | Q                                    |
| 3.007  |   | 30   | 73 m  | ~~                                       | ~~~   |   |   | 1350   |  |   | ~~  | ~~~  | ~~  | 1370   | mc n c   | ·m ~ m   | <b>C</b> 1                           |
| AGGA<br>E                                      | AGCC<br>A   |  | LAT   | GAC<br>T                                 | P   |   |   | CGATAT<br>D I  |  |   |   |  |   | L V  |  | rgi<br>V   |                                      |
| E  |   | 390  | м   | 1  | P   | K   | D   | 1410   | K  | 1.  | ٧   | G  | E   | 1430   | •  | ٧  | _                                    |
| <b>ምምር</b> ር                                   |   | -  | ጉ አ ጥ   | <b>ሮ</b> ልጥ                              | CCT   | COT   | CCT   | AGAGGT   | ייירר  | מסמי  | ቦልጥ                                       | سس   | CAC   |  | CCCT   | יראר   | THC                                  |
| G  | A   |  | I   | I  |   |   |   | E V  |  |   |   |  |   | M G  |  | T  | R                                    |
| ·  |   | 50   | _   | _  | _   | -   | •   | 1470   | •  | -   | -   | •  | •   | 1490   | •  | •  | •                                    |
| GCTT   |   |  |   |  |   |   |   |  |  |   |   |  |   |  |  |  |                                      |
|  | CTTI  |  | ACA   | GAC                                      | CAT   | CCT   | TGGG  | GGGCCC   | ATT  | 'CCA  | TGT                                       | CCT  | CAT   | CATCA  | CCTA   | TGC  | CŤ.                                  |
| F  | CTTI<br>F   | 'GG  |   |  |   |   |   | GGGCCC<br>G P  |  | CCA<br>H  |   |  |   | CATCA<br>I T   |  | TGC<br>A   |                                      |
|  | F   | 'GG  | ACA<br>Q  |  | CAT   |   |   |  |  |   |   |  |   |  |  |  |                                      |
| F  | F<br>15   | GGZ  | Q   | T  | I   | L   | G   | G P  | F  | Н   | V   | L  | I   | I T<br>1550  | Y  | . <b>A</b>   | F                                    |
| F  | F<br>15<br>GGTG   | GG<br>G<br>10<br>CT  | Q   | T  | I<br>CAT  | L<br>GGT  | G<br>GAT  | G P<br>1530  | F<br>CAT   | Н   | V<br>TGC                                  | L  | I<br>CGG  | I T<br>1550<br>GGAGG   | Y<br>TGG1  | . <b>A</b>   | F                                    |
| F<br>TCA1                                      | F<br>15<br>GGTG<br>V  | GG<br>G<br>10<br>CT  | G<br>Q  | T<br>Gac                                 | I<br>CAT  | L<br>GGT  | G<br>GAT  | G P<br>1530<br>GCGGCT  | F<br>CAT   | H<br>CAG  | V<br>TGC                                  | L<br>CAG                                     | I<br>CGG  | I T<br>1550<br>GGAGG   | Y<br>TGG1  | A<br>ACC   | F<br>CA                              |
| F<br>TCAT<br>M                                 | F<br>15<br>GGTG<br>V<br>15  | GG<br>G<br>10<br>CTC<br>L<br>70  | V<br>SGTN   | T<br>GAC<br>T                            | I<br>CAT<br>M   | L<br>GGT<br>V   | G<br>GAT<br>M   | G P<br>1530<br>GCGGCT<br>R L   | F<br>CAT<br>I  | H<br>CAG<br>S   | V<br>TGC<br>A                             | L<br>CAG<br>S                                | I<br>GGG  | I T<br>1550<br>GGAGG<br>E V<br>1610  | Y<br>TGGI<br>V                                   | A<br>ACC<br>P  | F<br>CA<br>M                         |
| F<br>TCAT<br>M                                 | F<br>15<br>GGTG<br>V<br>15<br>CTTT  | GGI<br>G<br>SCTC<br>L<br>70  | Q<br>SGTV<br>V<br>ACT                                     | T<br>GAC<br>T<br>CGT                     | i<br>Cat<br>M<br>GCT                                    | L<br>GGT<br>V<br>GGG  | G<br>GAT<br>M<br>CTG  | G P<br>1530<br>GCGGCT<br>R L<br>1590   | F<br>CAT<br>I  | H<br>CAG<br>S<br>CAT                                  | V<br>TGC<br>A<br>GTA                      | L<br>CAG<br>S<br>CTT                         | I<br>G<br>G   | I T<br>1550<br>GGGAGG<br>E V<br>1610<br>CCCGAG   | Y<br>TGGT<br>V<br>GATT                           | A<br>P<br>P  | F<br>CA<br>M                         |
| TCAT   | F<br>15<br>GGTG<br>V<br>15<br>CTTT  | GGI<br>G<br>SCTC<br>L<br>70  | Q<br>SGTV<br>V<br>ACT                                     | T<br>GAC<br>T<br>CGT                     | i<br>Cat<br>M<br>GCT                                    | L<br>GGT<br>V<br>GGG  | G<br>GAT<br>M<br>CTG  | G P<br>1530<br>GCGGCT<br>R L<br>1590<br>GTGCAA   | F<br>CAT<br>I  | H<br>CAG<br>S<br>CAT                                  | V<br>TGC<br>A<br>GTA                      | L<br>CAG<br>S<br>CTT                         | I<br>G<br>G   | I T<br>1550<br>GGGAGG<br>E V<br>1610<br>CCCGAG   | Y<br>TGGT<br>V<br>GATT                           | A<br>P<br>P  | F<br>CA<br>M                         |
| TCAT<br>M<br>TGTC                              | F<br>15<br>CGGTG<br>V<br>15<br>CCTTT<br>F   | GG<br>G<br>G<br>GCT<br>F<br>GC<br>A<br>GC  | Q<br>SGTV<br>V<br>ACT<br>L                                | T<br>GAC<br>T<br>CGT<br>V                | CAT<br>M<br>GCT<br>L                                    | L<br>GGT<br>V<br>GGG  | gate<br>m<br>ctge<br>w  | G P<br>1530<br>GCGGCT<br>R L<br>1590<br>GTGCAA<br>C N  | F<br>CAT<br>I<br>CGT<br>V  | H<br>CAG<br>S<br>CAT<br>M                             | V<br>TGC<br>A<br>GTA<br>Y                 | L<br>CAG<br>S<br>CTT                         | I<br>G<br>G<br>CGC<br>A   | I T<br>1550<br>EGGAGG<br>E V<br>1610<br>CCCGAG<br>R G<br>1670  | Y<br>TGGT<br>V<br>GATT<br>F                      | A<br>P<br>CCCA   | F<br>CA<br>M<br>GA<br>M              |
| TCAT M TGTC                                    | F<br>15<br>CGGTG<br>V<br>15<br>CCTTT<br>F   | GGI<br>G<br>G<br>GCT<br>GCI<br>A<br>GCI<br>A   | Q<br>SGT<br>V<br>ACT<br>L                                 | T<br>GAC<br>T<br>CGT<br>V                | CAT<br>M<br>GCT<br>L<br>CAT                             | EGT<br>V<br>GGG<br>G  | GATO<br>M<br>CTGO<br>W<br>GATO  | G P<br>1530<br>GCGGCT<br>R L<br>1590<br>GTGCAA<br>C N<br>1650  | F<br>CAT<br>I<br>CGT<br>V  | H<br>CAG<br>S<br>CAT<br>M                             | V<br>TGC<br>A<br>GTA<br>Y                 | CAG<br>S<br>CTT<br>F                         | CGC<br>A  | I T<br>1550<br>GGGAGG<br>E V<br>1610<br>CCCGAG<br>R G<br>1670  | Y<br>TGGI<br>V<br>GATI<br>F                      | A<br>P<br>P<br>CCA<br>Q  | ECA<br>M<br>GA<br>M                  |
| TCAT M TGTC                                    | F 15 CGGTG V 15 CCTTT F 16 CAGGG  | GGI<br>G<br>G<br>GCT<br>GCI<br>A<br>GCI<br>A   | Q<br>SGT<br>V<br>ACT<br>L                                 | T<br>GAC<br>T<br>CGT<br>V                | CAT<br>M<br>GCT<br>L<br>CAT                             | EGT<br>V<br>GGG<br>G  | GATO<br>M<br>CTGO<br>W<br>GATO  | G P 1530 GCGGCT R L 1590 GTGCAA C N 1650   | F<br>CAT<br>I<br>CGT<br>V  | H<br>CAG<br>S<br>CAT<br>M                             | V<br>TGC<br>A<br>GTA<br>Y                 | CAG<br>S<br>CTT<br>F                         | CGC<br>A  | I T<br>1550<br>GGGAGG<br>E V<br>1610<br>CCCGAG<br>R G<br>1670  | Y<br>TGGI<br>V<br>GATI<br>F                      | A<br>P<br>P<br>CCA<br>Q  | ECA<br>M<br>GA<br>M                  |
| TCAT M TGTC S TGCT                             | F 15 GGTG V 15 CCTTT F 16 CAGGC   | GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG   | Q<br>SGTV<br>V<br>ACT<br>L<br>CTTC<br>F                   | T<br>GAC<br>T<br>V<br>CAC<br>T           | CAT<br>M<br>GCT<br>L<br>CAT<br>I                        | L<br>CGT<br>V<br>GGG<br>G<br>CAT<br>M   | GATC<br>M<br>CTGC<br>W<br>GATT  | G P 1530 GCGGCT R L 1590 GTGCAA C N 1650 FCAGAA Q K 1710   | F<br>CAT<br>I<br>CGT<br>V<br>CAT<br>M  | H<br>CAG<br>S<br>CAT<br>M<br>GAT<br>I                 | V<br>TGC<br>A<br>GTA<br>Y<br>TTT<br>F     | L<br>CAG<br>S<br>CTT<br>F<br>TGG<br>G        | I GCGGG A D CTA   | I T 1550 GGAGG E V 1610 CCCAG R G 1670 ACCTGA L M 1730 ATATCA  | TGGT<br>V<br>GATT<br>F<br>TGCG<br>R              | A P CCA  | CA<br>M<br>GA<br>M<br>CT<br>C        |
| TCAT M TGTC S TGCT                             | F 15 GGTG V 15 CCTTT F 16 CAGGC   | GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG   | Q<br>SGTV<br>V<br>ACT<br>L<br>CTTC<br>F                   | T<br>GAC<br>T<br>V<br>CAC<br>T           | CAT<br>M<br>GCT<br>L<br>CAT<br>I                        | L<br>CGT<br>V<br>GGG<br>G<br>CAT<br>M   | GATC<br>M<br>CTGC<br>W<br>GATT  | G P 1530 GCGGCT R L 1590 GTGCAA C N 1650 FCAGAA Q K 1710   | F<br>CAT<br>I<br>CGT<br>V<br>CAT<br>M  | H<br>CAG<br>S<br>CAT<br>M<br>GAT<br>I                 | V<br>TGC<br>A<br>GTA<br>Y<br>TTT<br>F     | L<br>CAG<br>S<br>CTT<br>F<br>TGG<br>G        | I GCGGG A D CTA   | I T 1550 GGAGG E V 1610 CCCAG R G 1670 ACCTGA L M 1730 ATATCA  | TGGT<br>V<br>GATT<br>F<br>TGCG<br>R              | A P CCA  | CA<br>M<br>GA<br>M<br>CT<br>C        |
| TCAT M TGTC S TGCT L                           | F 15 CGTTT F 16 CAGGC G 16 GGCTG L  | PGGI<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G   | Q SGT V ACT L CTT F SGC A                                 | T<br>GAC<br>T<br>CGT<br>V<br>CAC<br>T    | I CAT M GCT L CAT I GGT V                               | L<br>GGT<br>V<br>GGG<br>G<br>CAT<br>M   | G GATE W GATE I CCTCL   | G P<br>1530<br>GCGGCT<br>R L<br>1590<br>GTGCAA<br>C N<br>1650<br>PCAGAA<br>Q K<br>1710<br>GGGCTT<br>G F<br>1770  | F CAT I  | H CCAG S CCAT M CGAT I TTCC                           | V TGC A GTA Y TTTT F AGC A                | L<br>CAG<br>S<br>CTT<br>F<br>TGG<br>G<br>CTT | I CGG G A CGA D CTA   | I T<br>1550<br>EGGAGG<br>E V<br>1610<br>CCCGAG<br>R G<br>1670<br>ACCTGA<br>L M<br>1730<br>ATATCA<br>I I<br>1790                  | Y<br>TGGI<br>V<br>GATI<br>F<br>TGCG<br>R<br>TCTI | A PACCE P CCE  | F CCA M GA M CCT C GA T              |
| TCAT M TGTC S TGCT L                           | F 15 CGGTG V 15 CCTTT F 16 CAGGC G 16 GGCTG L 17  | PGGI<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G   | Q<br>SGTV<br>V<br>ACTT<br>E<br>TTT<br>F<br>SGC'<br>A      | T GAC T CAC T TGT V                      | I CAT M GCT L CAT I GGT V                               | L<br>GGGT<br>V<br>GGGG<br>G<br>CAT<br>M<br>CAT<br>I   | G GATE M CTGG W GATE I CCTG L CCAG  | G P 1530 GCGGCT R L 1590 GTGCAA C N 1650 FCAGAA Q K 1710 GGGCTT G F 1770   | F CAT I CGAT M TGC A CGA   | H CCAG S CCAT M CGAT I CTTC S CCTA                    | V TGC A GTA Y TTTT F AGC A                | L CAG  | I CGG A CGA P CTA Y   | I T<br>1550<br>EGGAGG<br>E V<br>1610<br>CCCGAG<br>R G<br>1670<br>ACCTGA<br>L M<br>1730<br>ATATCA<br>I I<br>1790                  | Y TGGI V GATI F TGCG R TCTI F                    | A PACCE P CCE  | F CCA M GA M CCT C GA T              |
| TCAT M TGTC S TGCT L                           | F 15 CGGTG V 15 CCTTT F 16 CAGGC G 16 GGCTG L 17  | PGGI<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G   | Q<br>SGTV<br>V<br>ACTT<br>E<br>TTT<br>F<br>SGC'<br>A      | T GAC T CAC T TGT V                      | I CAT M GCT L CAT I GGT V                               | L<br>GGGT<br>V<br>GGGG<br>G<br>CAT<br>M<br>CAT<br>I   | G GATE M CTGG W GATE I CCTG L CCAG  | G P<br>1530<br>GCGGCT<br>R L<br>1590<br>GTGCAA<br>C N<br>1650<br>FCAGAA<br>Q K<br>1710<br>GGGCTT<br>G F<br>1770<br>CTTCTA<br>F Y   | F CAT I CGAT M TGC A CGA   | H CCAG S CCAT M CGAT I CTTC S CCTA                    | V TGC A GTA Y TTTT F AGC A                | L CAG  | I CGG A CGA P CTA Y   | I T<br>1550<br>EGGAGG<br>E V<br>1610<br>CCCGAG<br>R G<br>1670<br>LCCTGA<br>L M<br>1730<br>VTATCA<br>I I<br>1790<br>CCCTGT<br>L F | Y TGGI V GATI F TGCG R TCTI F                    | A PACCE P CCE  | CA M GA M CT C C                     |
| F TCAT M TGTC S TGCT L GCTC W CAGA             | F 155 CGGTG V 155 CCTTT F 166 G 166 GGCTG L 177 GGGAC D 18  | FGGZ<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G<br>G   | Q<br>CGGT<br>V<br>ACT<br>F<br>CGGC<br>A<br>CGAC<br>E      | T GAC T CGT V CAC T TGT V                | I CAT' M GCT' I GGT' V GCT' L                           | L<br>GGGT<br>V<br>GGG<br>G<br>CAT<br>M<br>CAT<br>I  | G GATC M CTG W GATT I CCTC L CCAC   | G P<br>1530<br>GCGGCT<br>R L<br>1590<br>GTGCAA<br>C N<br>1650<br>PCAGAA<br>Q K<br>1710<br>GGGCTT<br>G F<br>1770<br>CTTCTA<br>F Y<br>1830   | F CAT I LCGT W LGAT M LCGA D   | H CAG S CAT M CGAT I CTTC S CTA                       | V TGC A GTA Y TTT F AGC A CCC             | L CAG S CTT F TGG G CTT F CAT M              | I<br>CGGG<br>A<br>CGA<br>D<br>CTA<br>Y  | I T 1550 GGAGG E V 1610 CCCAG R G 1670 LCCTGA L M 1730 LTATCA I I 1790 CCCTGT L F  | Y TGGT V GATT F TGCG R TCTT F                    | A PACCA P CCCA CCCA CCCA CCCA CCCA CCCA  | F CA M GA M CT C C GA T CT F         |
| F TCAT M TGTC S TGCT L GCTC W CAGA E TCGA      | F 155 CGCTGT V 155 CGCTTT F 166 CGCTG L 176 CGGCCTG L 186 CGCCTG D 186 CGCTG  | PGGZ<br>G<br>G<br>GCTC<br>T<br>T<br>GCCC<br>P<br>GCCC<br>P<br>GCCC<br>P<br>GCCC<br>P<br>GCCC<br>P<br>GCCC<br>P<br>GCCC<br>P<br>GCCC<br>P<br>GCCC<br>P<br>GCCC<br>P<br>GCCC<br>P<br>GCCC<br>P<br>GCCC<br>P<br>GCCC<br>T<br>T<br>T<br>T<br>T<br>T<br>T<br>T<br>T<br>T<br>T<br>T<br>T<br>T<br>T<br>T<br>T   | Q SGT V ACT F SGC A CGA E                                 | T GAC T CGT V CAC T TGT V GGA E          | I CAT M GCT I GGT V GCT L CAT CAT CAT CAT               | L<br>GGGT<br>V<br>GGG<br>G<br>CAT<br>I<br>CAT<br>I  | G GATO M CTGG W GATO L CCAC H CCAC  | G P<br>1530<br>GCGGCT<br>R L<br>1590<br>GTGCAA<br>C N<br>1650<br>FCAGAA<br>Q K<br>1710<br>GGGCTT<br>G F<br>1770<br>CTTCTA<br>F Y<br>1830   | F CAT I I CGT V GAT M TGC A CGA D  | H CAG S CAT M GAT I TTC S CTA Y                       | V TGC A GTA Y TTT F AGC A CCC P           | L CAGE S CTT F G CTT F CAT M CAA             | I CGG A CGA Y CGG A CGG A   | I T 1550 GGAGG E V 1610 CCCAG R G 1670 LCCTGA L M 1730 LTATCA I I 1790 CCCTGT L F 1850   | Y TGGI V GATT F TGCG R TCTT F TCAG               | A PACCE P CCCE P CCCE T CCC T C | F CA M GA M CT C C GA T CT F CCA     |
| F TCAT M TGTC S TGCT L GCTC W CAGA             | F 155 CGCTCTT F 166 CGCCTCTT L 177 CGGAC D 188 CGCCTC L   | CCCCPP10   | Q SGT V ACT E SGC A CGA E CCT CCT                         | T GAC T CGT V CAC T TGT V GGA E          | I CAT M GCT I GGT V GCT L CAT CAT CAT CAT               | L<br>GGGT<br>V<br>GGG<br>G<br>CAT<br>I<br>CAT<br>I  | G GATO M CTGG W GATO L CCAC H CCAC  | G P<br>1530<br>GCGGCT<br>R L<br>1590<br>GTGCAA<br>C N<br>1650<br>PCAGAA<br>Q K<br>1710<br>GGGCTT<br>G F<br>1770<br>CTTCTA<br>F Y<br>1830<br>FGGCCC<br>G P                          | F CAT I CGT V AGAT M TGC A CGA D   | H CAG S CAT M GAT I TTC S CTA Y                       | V TGC A GTA Y TTT F AGC A CCC P           | L CAGE S CTT F G CTT F CAT M CAA             | I CGG A CGA Y CGG A CGG A   | I T 1550 GGAGG E V 1610 CCCAG R G 1670 LCCTGA L M 1730 ITATCA I I 1790 CCCTGT L F 1850 GGACC D L                                 | Y TGGI V GATT F TGCG R TCTT F TCAG               | A PACCE P CCCE P CCCE T CCC T CCC T CCC T CCC T CCC T C             | F CA M GA M CT C C GA T CT F CCA     |
| TCAT M TGTC S TGCT L GCTG W CAGA E TCGA        | F 155 CGGTG V 155 CGGTGT F 166 CGGGTG G 177 CGGAC D 18 CGGTG L 18 | GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG   | Q SGTV V ACTO E F SGCC A CGAC E CCT.                      | T GAC T CGT V CAC T GGA E TAC            | I CAT' M GCT' L CAT' L CAT' I                           | L<br>GGGG<br>G<br>CAT<br>I<br>AGG<br>G<br>CAT<br>I  | G GATCOM W GATCOM L CCCACA R CCCACA D   | G P<br>1530<br>GCGGCT<br>R L<br>1590<br>GTGCAA<br>C N<br>1650<br>PCAGAA<br>Q K<br>1710<br>GGGCTT<br>G F<br>1770<br>CTTCTA<br>F Y<br>1830<br>FGGCCC<br>G P<br>1890                  | F CAT I I CGAT M A CGA D A A A A A A A A A A A A A A A A A A   | H CCAG S CCAT M GGAT I TTC S CCTA Y CCAA N            | V TGC A GTA Y TTTT F AGC A CCC P CTA      | L CAG S CTT F TGG CTT F CAT M CAA            | I CGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG   | I T 1550 GGAGG E V 1610 CCCAG R G 1670 LCCTGA L M 1730 LTATCA I I 1790 CCCTGT L F 1850 CGGACC D L 1910                           | Y TGGI V GATI F TGCG R TCTI F TCAG               | A PACCE Q Q FACE T CCTT F  | GA M CT C GA T CT F CA M             |
| TCAT M TGTC S TGCT L GCTC W CAGA E TCGA        | F 155 CGGTG V 155 CGGTG V 155 CGGTG G 166 CGGGTG L 177 CGGAC D 18 CGGTG L 18 CGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG   | CCC P 10 CCC P 10 CCC P 17 TTC | Q SGTV V ACTO F F CGAC E CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC | T GAC T CGT V CAC T GGA E GGA T CTA      | I CAT' M GCT L CAT' V GCT L CAT' I TGC                  | L GGT V GGG G CAT I CAT | G GATON M GATON I COMMENT | G P 1530 GCGGCT R L 1590 GTGCAA C N 1650 FCAGAA Q K 1710 GGGCTT G F 1770 CTTCTA F Y 1830 FGGCCC G P 1890 FGCCAT  | F CAT I CGT V CGAT A CGAA D CGAA A CGAA  | H CCAG S CCAT M CGAT I CTAC S CTA Y CCAA N            | V TGC A GTA Y TTT F AGC A CCC P CTA Y CAC | L CAG S CTT F TGG G CTT F CAT M CAA          | I CGG A CGT | I T 1550 GGAGG E V 1610 CCCAG R G 1670 LCCTGA L M 1730 LTATCA I I 1790 CCCTGT L F 1850 CGGACC D L 1910 CCATGC                    | Y TGGT V  GATT F TCTT F TCAG S TCCC P            | A PACCE P CCCE F CCCE T CCCE F ACCE ACCE   | GA M CT C GA T CT F CA M CC C        |
| TCAT M TGTC S TGCT L GCTC W CAGA E TCGA        | F 155 CGGTG V 155 CGGTG V 155 CGGTG G 166 CGGGTG L 177 CGGAC D 18 GGGTG L 18 CGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG   | CCCC P 10 CCCC P | Q SGTV V ACTO F F CGAC E CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC | T GAC T CGT V CAC T GGA E GGA T CTA      | I CAT' M GCT L CAT' V GCT L CAT' I TGC                  | L GGT V GGG G CAT I CAT | G GATON M GATON I COMMENT | G P<br>1530<br>GCGGCT<br>R L<br>1590<br>GTGCAA<br>C N<br>1650<br>PCAGAA<br>Q K<br>1710<br>GGGCTT<br>G F<br>1770<br>CTTCTA<br>F Y<br>1830<br>FGGCCC<br>G P<br>1890<br>FGCCAT<br>A I | F CAT I CGT V CGAT A CGAA D CGAA A CGAA  | H CCAG S CCAT M CGAT I CTAC S CTA Y CCAA N            | V TGC A GTA Y TTT F AGC A CCC P CTA Y CAC | L CAG S CTT F TGG G CTT F CAT M CAA          | I CGG A CGT | I T 1550 GGAGG E V 1610 CCCAG R G 1670 LCCTGA L M 1730 CCTGA I I 1790 CCTGT L F 1850 CGGACC D L 1910 CCATGC M L                  | Y TGGT V  GATT F TCTT F TCAG S TCCC P            | A PACCE P CCCE F CCCE T CCCE F ACCE ACCE   | GA M CT C GA T CT F CA M CC C        |
| TCAT M TGTC S TGCT L GCTG W CAGA E TCGA E TGTA | F 155 CGGTG V 155 CGGTG V 155 CGGTG G 166 CGGGGG L 177 CGGAC D 18 GGGGG L 18 CGGGG S 19   | CCCC P 10 CCCC P | Q SGT V ACT F EGG A CGAG E CCT L CACG                     | T GAC T CGT V CAC T GGAC T T GTAC T CTAC | I CAT M GCT I CAT I I I I I I I I I I I I I I I I I I I | L<br>GGT<br>V<br>GGG<br>G<br>CAT<br>I<br>CAT<br>I<br>CAT<br>I<br>FGC  | G GAT' M CTG' W GAT' I CCTG' H CGA' D CTT':   | G P 1530 GCGGCT R L 1590 GTGCAA C N 1650 FCAGAA Q K 1710 GGGCTT G F 1770 CTTCTA F Y 1830 FGGCCC G P 1890 FGCCAT  | F CAT I LCGT V LGAT A LCGA A LCGA A LCGA I L | H TCAG S TCAT M TGAT I TCTTC S TCTA Y TCCAA N TCCGC A | V TGC A GTA Y TTT F AGC A CCC P CTA Y CAC | L CAG S CTT F TGG G CTT F CAT M CAA N ACT L  | I CGG A CGT A CGT V CGCT L  | I T 1550 GGAGG E V 1610 CCCAGG R G 1670 LCCTGA L M 1730 CCTGA I I 1790 CCTGT L F 1850 CGGACC D L 1910 CCATGC M L 1970            | Y TGGT V GATT F TGCG R TCTT F TCAG               | A P CCCP Q CCCP F CCCP T CCCP T CCCP L   | F CA M GA M CT C GA T CT F CA M CC L |

Fig. 8 / continui on 2

I A M M G D T H W R V A H E R D E L W R

1990 2010 2030

GGGCCCAGATTGTGGCCACCACGGTGATGCTGGAGCGGAAGCTGCCTCGCTGCCTGTGGC
A Q I V A T T V M L E R K L P R C L W P

2050 2070 2090

CTCGCTCCGGGATCTGCGGACGGGATATGGCCTGGGAGACCGCTGGTTCCTGCGGGTGG

R S G I C G R E Y G L G D R W F L R V E
2110 2130 2150

ARGACAGGCAAGATCTCAACCGCAGCCGATCCAACGCTACGCACAGGCCTTCCACACCC

ARGACAGGCAAGATCTCAACCGCAGCGATCCAACGCTACGCACAGGCCTTCCACACCC

D R Q D L N R Q R I Q R Y A Q A F H T R

2170 2190 2210

GGGGCTCTGAGGATTTGGACAAAGACTCAGTGGAAAAACTAGAGCTGGGCTGTCCCTTCA
G S E D L D K D S V E K L E L G C P F S
2230 2250 2270

GCCCCCACCTGTCCCTTCCTATGCCCTCAGTGTCTCGAAGTACCTCCCGCAGCAGTGCCA
P H L S L P M P S V S R S T S R S S A N
2290 2310 2330

ATTGGGAAAGGCTTCGGCAAGGGACCCTGAGGAGAGACCTGCGTGGGATAATCAACAGGG W E R L R Q G T L R R D L R G I I N R G 2350 2370 2390

GTCTGGAGGACGGGGAGACTGGGAATATCAGATCTGA
L E D G E S W E Y Q I \*

MGLSLPKEKGLILCLWSKFCRWFQRRESWAQSRDEQNLLQQKRIWESPLLLAAKDNDVQALNKILKYEDCKVHQRGAMGETALHIA ALYDNLEAAMVLMEAAPELVFEPMTSELYEGGTALHIAVVNQNMNLVRALLARRASVSARATGTAFRRSPCNLIYFGBHPLSFAAC VNSEEIVRLLIEHGADTRAQDSLGNTVLHILTLQPNKTFACQMYNLLLSYDRHGDHLQPLDLVPNHQGLTPFKLAGVBGNTVMFQB LMQKRKHTQWTYGPLTSTLYDLTEIDSSGDEQSLLELIITTKKREARQILDQTPVKELVSLKWKRYGRPYFCMLGAIYLLYIICFT MCCIYRPLKPRTNNRTSPRDNTLLQQKLLQEAYMTPKDDIRLVGELVTVIGAIIILLVEVPDIFRMGVTRFFGQTILGGPFHVLII TYAFMVLVTMVMRLISASGEVVPMSFALVLGWCNVMYFARGFQMLGPFTIMIQKMIFGDLMRFCWLMAVVILGFASAFYIIFQTED PEELGHFYDYPMALFSTFELFLTIIDGPANYNVDLPFMYSITYAAFAIIATLIMLNLLIAMMGDTHWRVAHERDELWRAQIVATTV MLERKLPRCLWFRSGICGREYGLGDRWFLRVEDRQDLNRQRIQRYAQAFHTRGSEDLDKDSVEKLELGCPFSPHLSLPMPSVSRST SRSSANWERLRQGTLRRDLRGIINRGLEDGESWEYQI

B)

CCTCTACAGGGAGACGGTGGGCCGGCCCTTGGGGGGGCTGATGTGGCCCCAAGGCTGAGTCCCGTCAGGGTCTGGCCTCGGCCTCA GGCCCCCAAGGAGCCGGCCCTACACCCCATGGGTTTGTCACTGCCCAAGGAGAAGGGCTAATTCTCTGCCTATGGAGCAAGTTCT GCAGATGGTTCCAGAGACGGGAGTCCTGGGCCCAGAGCCGAGATGAGCAGAACCTGCTGCAGCAGAAGAGGATCTGGGAGTCTCCT CTCCTTCTAGCTGCCAAAGATAATGATGTCCAGGCCCTGAACAAGTTGCTCAAGTATGAGGATTGCAAGGTGCACCAGAGAGGAGG CATGGGGGAAACAGCGCTACACATAGCAGCCCTCTATGACAACCTGGAGGCCGCCATGGTGCTGATGGAGGCTGCCCCGGAGCTGG TCTTTGAGCCCATGACATCTGAGCTCTATGAGGGTCAGACTGCACTGCACATCGCTGTTGTGAACCAGAACATGAACCTGGTGCGA GCCCTGCTTGCCCGCAGGGCCAGTGTCTCTGCCAGAGCCACAGGCACTCCCTTCCGCCGTAGTCCCCGCAACCTCATCTACTTTGG GGAGCACCCTTTGTCCTTTGCTGCTGTGTGAACAGTGAGGAGATCGTGCGGCTGCTCATTGAGCATGGAGCTGACATCCGGGCCC TGCAGAAGCGGAAGCACACCCAGTGGACGTATGGACCACTGACCTCGACTCTCTATGACCTCACAGAGATCGACTCCTCAGGGGAT GAGCAGTCCCTGCTGGAACTTATCATCACCACCAAGAAGCGGGAGGCTCGCCAGATCCTGGACCAGACGCCGGTGAAGGAGCTGGT GAGCCTCAAGTGGAAGCGGTACCGGCCGTACTTCTGCATGCTGGGTGCCATATATCTGCTGTACATCATCTGCTTCACCATGT GCTGCATCTACCGCCCCCTCAAGCCCAGGACCAATAACCGCACAAGCCCCCGGGACAACACCCTCTTACAGCAGAAGCTACTTCAG GAAGCCTACGTGACCCCTAAGGACGATATCCGGCTGGTCGGGGGAGCTGGTGACTGTCATTGGGGCTATCATCATCCTGCTGGTAGA GGTTCCAGACATCTTCAGAATGGGGGTCACTCGCTTCTTTGGACAGACCATCCTTGGGGGCCCATTCCATGTCCTCATCATCACCT TCATCCTGGGCTTTGCTTCAGCCTTCTATATCATCTTCCAGACAGGAGCCCGAGGAGCTAGGCCACTTCTACGACTACCCCATG GCCCTGTTCAGCACCTTCGAGCTGGTCCTTACCATCATCGATGGCCCAGCCAACTACAACGTGGACCTGCCCTTCATGTACAGCAT CCCATGAGCGGGATGAGCTGTGGAGGGCCCAGATTGTGGCCACCACGGTGATGCTGGAGCGGAAGCTGCCTCGCTGCCTGTGGCCT CGCTCCGGGATCTGCGGACGGGAGTATGGCCTGGGGGACCGCTGGTTCCTGCGGGTGGAAGACAGGCAAGATCTCAACCGGCAGCG

c.)

CAAACTCACAGCCCTCTCCAAACTGGCTGGGGGCTGCTGGGAGACTCCCAAGGAACTCGTCAGGAAGGCAGGAGACACGGAGACACGGGA CCTCTACAGGGAGACGGTGGGCCGCCCTTGGGGGGGGCTGATGTGGCCCCAAGGCTGAGTCCCGTCAGGGTCTGGCCTCGGCCTCA GGCCCCCAAGGAGCOGGCCCTACACCCCATGGGTTTGTCACTGCCCAAGGAGAAAGGGCTAATTCTCTGCCTATGGAGCAAGTTCT CTCCTTCTAGCTGCCAAAGATAATGATGTCCAGGCCCTGAACAAGTTGCTCAAGTATGAGGATTGCAAGGTGCACCAGAGAGGAGC CATGGGGAACAGCGCTACACATAGCAGCCCTCTATGACAACCTGGAGGCCGCCATGGTGCTGATGGAGGCTGCCCCGGAGCTGG TCTTTGAGCCCATGACATCTGAGCTCTATGAGGTCCTGACTGCCCATCACTTGAAGGCCTGCCCCCTGAAATGCCAGGGCCTAGAG AAGAGGAAGAGTGGGCAGCTGGATCCCCTGGGAATCCTGAACACCCGAGAGCTCCCTGTTCTCCATCCCAGGCTACCCCTGA TCAGACTGCACTGCACATCGCTGTTGTGAACCAGAACATGAACCTGGTGGGAGCCCTGCTTGCCCGCAGGGCCAGTGTCTCTGCCA GAGCCACAGGCACTGCCTTCCGCCGTAGTCCCTGCAACCTCATCTACTTTGGGGAGCACCCTTTGTCCTTTGCTGCCTGTGTGAAC AGATGTACAACCTGTTGCTGTCCTACGACAGACATGGGGACCACCTGCAGCCCTGGACCTCGTGCCCAATCACCACGGTCTCACC CCTTTCAAGCTGGCTGGAGTGGAGGGTAACACTGTGATGTTTCAGCACCTGATGCAGAAGCGGAAGCACCCAGTGGACGTATGG ACCACTGACCTCGACTCTCTATGACCTCACAGAGATCGACTCCTCAGGGGGTTGAGCAGTCCCTGCTGGAACTTATCATCACCACCA AGAAGCGGGAGGCTCGCCAGATCCTGGACCAGACGCCGGTGAAGGAGCTGGTGAGCCTCAAGTGGAAGCGGTACGGGCCGTAC TTCTGCATGCTGGGTGCCATATATCTGCTGTACATCATCTGCTTCACCATGTGCTGCATCTACCGCCCCCTCAAGCCCAGGACCAA TAACCGCACGAGCCCCCGGGACAACACCCCTCTTACAGCAGAAGCTACTTCAGGAAGCCTACATGACCCCTAAGGACGATATCCGGC TGGTCGGGAGCTGGTGACTGTCATTGGGGCTATCATCCTGCTGGTAGAGGTTCCAGACATCTTCAGAATGGGGGTCACTCGC  $\tt TTCTTTGGACAGACCATCCTTGGGGGGCCCATTCCATGTCCTCATCATCATCATCATGCTTCATGGTGCTGGTGATCCATGGTGATGCG$ GCTCATCAGTGCCAGCGGGGAGGTGGTACCCATGTCCTTTGCACTCGTGCTGCGTGCTACCTCATGTACTTCGCCCGAGGAT ATCCTGGGCTTTGCTTAGACAGAGGAGCCCCGAGGAGCTAGGCCACTTCTACGACTACCCCATGGCCCTGTTCAGCACCTTCGAGCT GGTCCTTACCATCATCGATGGCCCAGCCAACTACAACGTGGACCTGCCCTTCATGTACAGCATCACCTATGCCGCTTTGCCATCA AGGGCCCAGATTGTGGCCACCACGGTGATGCTGGACCGGAGCTGCCTCGCTGCCTGTGGCCTCGCTCCGGGATCTGCGGACGGGA GTATGGCCTGGGAGACCGCTGCTTCCTGCGGGTGGAAGACAGGCAAGATCTCAACCGCTAGGGGATCCAACGCTACGCACAGGCCT TCCACACCCGGGGCTCTGAGGATTTGGACAAAGACTCAGTGGAAAAACTAGAGCTGGGCTGTCCCTTCAGCCCCCACCTGTCCCTT CCTATGCCCTCAGTGTCTCGGAAGTACCTCCCGCAGCAGTGCCAATTGGGAAAGGCTTCGGCAAGGGACCCTGAGGAAGACCTGCG TGGGATAATCAACAGGGGTCTGGAGGACGGGGAGAGCTGGGAATATCAGATCTGACTGCGTGTTCTCACTTCGCTTCCTGGAACTT GCTCTCATTTTCCTGGGTGCATCAAACAAAACAAAAACCAAACACCCAGAGGTCTCATCTCCCAGGCCCCCAGGGAAAGAAGAGGAGT AGCATGAACGCCAAGGAATGTACGTTGAGAATCACTGCTCCAGGCCTGCATTACTCCTTCAGCTCTGGGGCAGAGGAAGCCCAGCC CAAGCACGGGCTGGCAGGGCGTGAGGAACTCTCCTGTGGCCTGCTCATCACCCTTCCGACAGGAGCACTGCATGTCAGAGCACTT Tararacaggccagcctgcttgggccctcggtctccaccccaggctcataagtggggagagagcccttcccagggcaccag CTGCASGGAAGTGCAGAGCTTGTGGAAAGCGTGTGAGTGAGGGAGACAGGAACGGCTCTGGGGGAAGTGGGGCTAGGTCTTG 

٥.

CAAACTCACAGCCCTCTCCAAACTGGCTGGGGGGGTGCTGGGAGACTCCCAAGGAACTCGTCAGGAAGGCAGGAGACAGGAACACGGA CCTCTACAGGGGAGACGGTGGGCCCCTTGGGGGGGGGCTGATGTGGCCCCAAGGCTGAGTCCCGTCAGGGTCTGGCCTCA GGCCCCCAAGGAGCCGGCCCTACACCCCATGGGTTTGTCACTGCCCAAGGAGAAAGGGCTAATTCTCTGCCTATGGAGCAAGTTCT GCAGATGGTTCCAGAACGGGAGTCCTGGGCCCAGAGCCGAGATGACCAGAACCTGCTGCAGCAGAAGAGGATCTGGGAGTCTCCT CTCCTTCTAGCTGCCAAAGATAATGATGTCCAGGCCCTGAACAAGTTGCTCAAGTATGAGGATTGCAAGGTGCACCAGAAGAGGACC CATGGGGGGAAACAGCGCTACACATAGCAGCCCTCTATGACAACCTGGAGGCCGCCATGGTGCTGATGAGGGCTGCCCCGGAGCTGG TCTTTGAGCCCATGACATCTGAGCTCTATGAGGGTCAGACTGCACTGCTCTTTGTGAACCAGAACATGAACCTGGTGCGA GCCCTGCTTGCCCCCAGGGCCAGTGTCTCTCCCCAGAGCCACAGGCCCCTTCCCCCGCAACCTCATCTTCTTTTGG Fig. 8 / continuation 4

GTGATGTTTCAGCACCTGATGCAGAAGCGGAAGCACCCAGTGGACGTATGGACCACTGACCTCGACTCTATGACCTCACAGA GATCGACTCCTCAGGGGATGAGCAGTCCCTGCTGGAACTTATCATCACCACCAAGAAGCGGGAGGCTCGCCAGATCCTGGACCAGA ATCATCTGCTTCACCATGTGCTGCATCTACCGCCCCCTCAAGCCCCAGGACCAATAACCGCACAAGCCCCCGGGACAACACCCCTCTTA CAGCAGAAGCTACTTCAGGAAGCCTACGTGACCCCTAAGGACGATATCCGGCTGGTCGGGGGGGCTGGTCATTGGGGCTATCATCATCCTGCTGGTAGAGGTTCCAGACATCTTCAGAATGGGGGTCACTCGCTTCTTTGGACAGACCATCCTTGGGGGCCCCATTC CATGTCCTCATCATCACCTATGCCTTCATGGTGCTGGTGACCATGGTGATGCGGCTCATCAGTGCCAGCGGGGAGGTGGTACCCAT GTCCTTTGCACTCGTGCTGGGCTGGTGCAACGTCATGTACTTCGCCCGAGGATTCCAGATGCTAGGCCCCCTTCACCATCATGATTC TTCCAGACAGAGGACCCGAGGAGCTAGGCCACTTCTACGACTACCCCATGGCCCTGTTCAGCACCTTCGAGCTGGTCCTTACCAT CATCGATGCCCAGCCAACTACAACGTGGACCTGCCCTTCATGTACAGCATCACCTATGCCGCCTTTGCCATCATCGCCACACTGC GGACCGCTGGTTCCTGCGGGTGGAAGACAGGCAAGATCTCAACCGGCAGCGGATCCAACGCTTACGCACAGGCCTTCCACACCCGGG GCTCTGAGGATTTGGACAAAGACTCAGTGGAAAAACTAGAGCTGGGCTGTCCCTTCAGCCCCACCTGTCCCTTCCTACGCCCTCA GTGTCTCGAAGTACCTCCCGCAGCAGTGCCAATTGGGAAAGGCTTCGGCAAGGGACCTGAGGAGAGACCTGCGTGGGATAATCAA Caggestttsgaggaggagasctsggaatatcagatctsactsctsttttcacttcscttcctsgaacttsctcattttc CTGGGTGCATCAAACAAAACAAAACCAAACCCAGAGGTCTCATCTCCCAGGCCCCCAGGGAGAAAGAGGAGTAGCATGAACGCC AAGGAATGTACGTTGAGAATCACTGCTCCAGGCCTGCATTACTCCTTCAGCTCTGGGGCAGAGGAAGCCCCAGCCCAAGCACGGGGC TGGCAGGGCGTGAGGAACTCTCCTGTGGCCTGCTCATCACCCTTCCGACAGGAGCACTGCATGTCAGAGCACTTTAAAAACAGGCC GCAGAGCTTGTGGAAAGCGTGTGAGTGAGGGAGACAGGAACGGCTCTGGGGGTGGGAAGTGGGGCTAGGTCTTGCCAACTCCATCT

E.)

CACACATEGGGCCTCCCAGGAGTGCCCAGGACCTCGTGCTGTTGGCCTCTGAATCTATCGTCTCCAATCCGCTGTCCCACAGAAGC CATATAACCCACCTCTCTGTAAATGCCAGGAGCCATGGGGGAAACAGCGCTACACATAGCAGCCCTCTATGACAACCTGGAGGCCC CCATGGTGCTGATGGAGGCTGCCCCCGGAGCTGGTCTTTGAGCCCCATGACATCTGAGCTCTATGGAGGGTTGAGGGCCCACGGGTCTG CCTACTCTTTTTSTCTTCTCTGTCTCCCTTCCGTGTCAGTCCCTGACTGCCCATCACTTGAACGCCTGCCCCTGAAATGCCAGGG GCCTAGAGAAGAGGAGGAGGAGCAGCAGCTGGATCCCCTGGGAATCCTGAACACCCGAGAGCTCCCTGTTCTCCATCCCAGGCT  $\tt CTGGGCCAGGTCAGACTGCACTGCACATCGCTGTTGTAACCAGAACATGAACCTGGTGCGAGGCCCTGCTTGCCCGCAGGGCCAGT$ GTCTCTGCCAGAGCCACAGGCACTGCCTTCCGCCGTAGTCCCTGCAACCTCATCTACTTTTGGGGAGCACCCTTTGTCCTTTGCTGC CTCTCTGAACAGTGAGGAGATCCTGCGGCTCCTCATTGAGCATGGAGCTGACATCCGGGCCCCAGGACTCCCTGGATGTACAACCTG TTGCTGTCCTACGACAGCATGGGGACCACCTGCAGCCCTGGACCTCGTGCCCAATCACCAGGGTCTCACCCCTTTCAAGCTGGC TGGAGTGGAGGGTAACACTGTGATGTTTCAGCACCTGATGCAGAAGCGGAAGCACACCCAGTGGACGTATGGACCACTGACCTCGA CTCTCTATGACCTCACAGAGATCGACTCCTCAGGGGATGAGCAGTCCCTGCTGGAACTTATCATCACCACCAAGAAGCGGGAGGCT CGCCAGATCCTGGACCAGACGCCGTTGAAGGAGCTGGTGAGCCTCAAGTGGAAGCGGTACGGCCGTACTTCTGCATGCTGGG TGCCATATATCTGCTGTACATCATCTGCTTCACCATGTGCTTCTACCGCCCCCTCAAGCCCAGGACCAATAACCGCACGAGCC CCCGGGACACACCCTCTTACAGCAGAAGCTACTTCAGGAAGCCTACATGACCCCTAAGGACGATATCCGGCTGGTCGGGGAGCTG GTGACTGTCATTGGGGCTATCATCATCCTGCTGGTAGAGGTTCCAGACATCTTCAGAATGGGGGTCACTCGCTTCTTTGGACAGAC CATCCTTGGGGGCCCATTCCATGTCCTCATCACCTATGCCTTCATGGTGCTGGTGACCATGGTGATGCGGCTCATCAGTGCCA GCGGGGGGGGTGCCCATGTCCTTTGCACTCGTGCTGGGCTGGTGCAACGTCATGTACTTCGCCCGAGGATTCCAGATGCTAGGC TTCAGCCTTCTATATCATCTTCCAGACAGAGGACCCCGAGGAGCTAGGCCACTTCTACGACTACCCCATGGCCCTGTTCAGCACCT  ${\tt TCGAGCTGGTCCTTACCATCATCGATGGCCCAGCCAACTACAACGTGGACCTGCCCTTCATGTACAGCATCACCTATGGTCCCTTT$ GACGGAGTATGGCTTGGGAGACCGCTGGTTCCTGCGGGTGGAAGACAGGCAAGATCTCAACCGGAGCGGATCCAACGCTACGCA  ${\tt CAGGCCTTCCACACCCGGGCCTCTGAGGATTTGGACAAAGACTCAGTGGAAAAACTAGAGCTGGGCTGTCCCTTCAGCCCCCACCT}$ GTCCCTTCCTATGCCCTCAGTGTCTCGAAGTACCTCCCGCAGCAGTGCCAATTGGGAAAGGCTTCGGCAAGGGACCCTGAGGAGAC ACCTGCGTGGGATAATCAACAGGGGTCTGGAGGACGGGGAGAGCTGGGAATATCAGATCTGACTGCGTGTTCTCACTTCGCTTCCT GGAACTTGCTCTCATTTTCCTGGGTGCATCAAACAAAACCAAAACCCAGAGGTCTCATCTCCCAGGCCCCAGGGAGAAA GAGGAGTAGCATGAACGCCAAGGAATGTACGTTGAGAATCACTGCTCCAGGCCTGCATTACTCCTTCAGCTCTGGGGCAGAGGAAG CCCAGCCCAAGCACGGGGCTGGCAGGGAACTCTCCTGTGGCCTCATCACCCTTCCGACAGGAGCACTGCATGTCAG AGCACTTTAAAAACAGGCCAGCCTGCTTGGGCCCTCGGTCTCCACCCCAGGGTCATAAGTGGGGAGAGAGCCCTTCCCAGGGCACC

19/40

Fig. 8 / continuation 5

Figure 9:

A.

|   |      | 10   |      | 30   |        |             |      |     |        |      |       |       | 50           |      |        |     |     |      |     |  |  |
|---|------|------|------|------|--------|-------------|------|-----|--------|------|-------|-------|--------------|------|--------|-----|-----|------|-----|--|--|
| CGGG                                    | GCC  | CTG  | GGC  | TGC: | AGG    | AGG         | TTG  | CGG | CGG    | CCG  | CGG   | CAG   | CAT          | GGT  | GGT    | GCC | GGA | GAA  | GG  |  |  |
|   |      |      |      |      |        |             |      |     |        |      |       |       | M            | ٧    | V      | ₽   | E   | K    | E   |  |  |
|   |      | 7Ó   |      |      |        |             |      |     | 90     |      |       |       |              |      | 11     | 0   |     |      |     |  |  |
| AGCA                                    | GAG  | CTG  | GAI  | CCC  | CAA    | GAT         | CTT  | CAA | GAA    | GAA  | GAC   | CTG   | CAC          | GAC  | GTT    | CAT | AGT | TGA  | CT  |  |  |
| Q                                       | S    | W    | 1    | P    | ĸ      | I           | F    | K   | K      | K    | T     | С     | T            | T    | F      | I   | V   | D    | s   |  |  |
|   |      | 130  |      |      |        |             |      | 1   | 50     |      |       |       |              |      | 17     | 0   |     |      |     |  |  |
| CCAC                                    | :AGI | ATCO | GGG  | AGG  | GAC    | CTT         | GTG  | CCA | GTG    | TGG  | GCG   | ccc   | CCG          | GÁC  | CGC    | CCA | CCC | CGC  | AG  |  |  |
| T                                       | D    | P    | G    | G    | т      | L           | С    | 0   |        | G    | R     | P     | R            | Т    | A      | н   | P   | A    | v   |  |  |
|   |      | 190  |      |      |        |             |      | .2  | 10     | _    |       | _     |              | _    | 23     | 0   | _   |      |     |  |  |
| TGGC                                    | CAT  | rgga | GGZ  | TGC  | CTT    | CGG         | GGC  |     |        | 'GGT | CAC   | CGT   | 'CTC         | CGA  |        | -   | ጥርር | מחמי | CA  |  |  |
| A                                       | М    | E    | D    | A    | F.     |             | A    |     | v      | v    | T     | V     | W            | D    |        | D   |     | H    |     |  |  |
|   |      | 250  | _    | ••   |        | ٠           | ••   |     | 70     | ٠    | •     | •     | ••           |      | 29     | _   | 41  | 44.  | •   |  |  |
| CCAC                                    | cci  |      | ccc  | יראר | CCA    | ጥርር         | ' ጥን | _   | _      | CCT  | CCD   | COURT | C N C        | ·ccc |        | -   | ccc | ראא  | CC  |  |  |
| T                                       | E    | K    | P    | T    | D.     | A           | Y    | G   |        |      |       |       | T            | .GGG | A<br>A |     | R   |      |     |  |  |
| 1                                       | -    | 310  | •    | 7    | D      | A           | 1    |     |        | L    | D     | F     | T            | G    |        |     | K   | K    | H   |  |  |
| 2020                                    |      |      |      |      |        |             |      | _   | 30<br> |      |       |       |              |      | 35<br> | -   |     |      |     |  |  |
| ACAG                                    |      |      |      |      |        |             |      |     |        |      |       |       |              |      |        |     |     |      |     |  |  |
| S                                       | N    | F    | L    | R    | L      | S           | D    | R   | _      | D    | P     | A     | A            | V    | _      | S   | L   | V    | . T |  |  |
|   |      | 370  |      |      |        |             |      | _   | 90     |      |       |       |              |      | 41     | -   |     |      |     |  |  |
| CACC                                    |      |      |      |      |        |             |      |     |        |      |       |       |              |      |        |     |     | GGG  | GG  |  |  |
| R                                       | T    | W    | G    | F    | R      | A           | P    | N   | _      | ٧    | V     | S     | V            | L    | _      | G   | S   | G    | G   |  |  |
|   |      | 430  |      |      |        |             |      | _   | 50     |      |       |       |              |      | 47     | -   |     |      |     |  |  |
| GCCC                                    |      |      | CCA  | GAC  |        | GCT         | GCA  | GGA | CCI    | 'GCI | GCG   | TCG   | TGG          | GCT  |        |     | GGC | TGC  | :cc |  |  |
| P                                       | V    |      | Q    | T    | W      | L           | Q    | D   | L      | L    | R     | R     | G            | L    | V      | R   | A   | A    | Q   |  |  |
|   |      | 490  |      |      |        |             |      | 5   | 10     |      |       |       |              |      | 53     | 0   |     |      |     |  |  |
| AGAG                                    | CAC  | CAGG | AGC  | CTG  | GAT    | TGT         | CAC  | TGG | GGG    | TCI  | GCA   | CAC   | GGG          | CAT  | CGG    | CCG | GCA | TGT  | TG. |  |  |
| S                                       | T    | G    | A    | W    | I      | V           | T    | Ģ   | G      | L    | H     | T     | G            | I    | G      | R   | H   | V    | G   |  |  |
|   |      | 550  |      |      |        |             |      | 5   | 70     |      |       |       |              |      | 59     | 0   |     |      |     |  |  |
| GTGI                                    | 'GGC | TGT  | ACG  | GGA  | CCA    | TCA         | GAT  | GGC | CAG    | CAC  | TGG   | GGG   | CAC          | CAA  | GGT    | GGT | GGC | CAT  | 'GG |  |  |
| V                                       | A    | V    | R    | D    | H      | Q           | M    | A   | S      | T    | G     | G     | T            | K    | V      | ٧   | A   | M    | G   |  |  |
|   | •    | 610  |      |      |        |             |      | 6   | 30     |      |       |       |              |      | 65     | 0   |     |      |     |  |  |
| GTGT                                    | 'GGC | ccc  | CTG  | GGG  | TGT    | GGT         | CCG  | GAA | TAG    | AGA  | CAC   | CCI   | CAT          | 'CAA | CCC    | CAA | GGG | CTC  | GT  |  |  |
| v                                       | A    | P    | W    | G    | V      | V           | R    | N   | R      | D    | T     | L     | I            | N    | P      | K   | G.  | s    | F   |  |  |
|   |      | 670  |      |      |        |             |      | 6   | 90     |      |       |       |              |      | 71     | 0   |     | •    |     |  |  |
| TCCC                                    | TGC  | GAG  | GTA  | CCG  | GTG    | GCG         | CGG  | TGA | CCC    | GGA  | .GGA  | CGG   | GGT          | CCA  | GTT    | TCC | CCT | GGA  | CT  |  |  |
| P                                       | A    | R    | Y    | R    | W      | R           | G    | D   | P      | E    | D     | G     | V            | Q    | F      | P   | L   | D    | Y   |  |  |
|   |      | 730  |      |      |        |             |      | 7.  | 50     |      |       |       |              |      | 77     | 0   |     |      |     |  |  |
| ACAA                                    | CT   | CTC  | GGC  | CTT  | CTT    | CCT         | GGT  | GGA | CGA    | .CGG | CAC   | ACA   | .CGG         | CTG  | CCT    | GGG | GGG | CGA  | GA  |  |  |
| N                                       | Y    | s    | Α    | F    | F      | L           | V    | D   | D      | G    | T     | н     | G            | С    | L      | G   | G   | E    | N   |  |  |
|   |      | 790  |      |      |        |             |      | 8   | 10     |      |       |       |              |      | 83     | 0   |     |      |     |  |  |
| ACCG                                    | CTI  | CCG  | CTT  | GCG  | CCT    | GGA         | GTC  | _   | -      | CTC  | ACA   | GCÀ   | GAA          | GAC  |        |     | GGG | AGG  | GA  |  |  |
|   |      | R    |      |      |        |             |      |     |        |      |       |       |              |      |        |     |     |      |     |  |  |
| • | •    | 850  | _    | ••   | -      | -           | •    |     | 70     | -    | ¥     | ×     | • •          | •    | 89     |     | •   | •    | •   |  |  |
| CTGG                                    | ከከጥ  |      | ግ አጥ | CCC  | arCar. | <b>ጉ</b> ጉጥ | CCT  | _   |        | ርካጥ  | ብር: አ | me e  | י.<br>יייביא | ሞርአ  |        |     | ششت | CAC  | cc  |  |  |
| G                                       |      |      |      |      |        |             |      |     |        |      |       | G     |              |      |        | M   |     |      |     |  |  |
| G                                       | 1    | D    | 1    | r    | ٧      | ינ          |      |     |        | +    | D     | G     | ט            | E    |        |     | L   | T    | М   |  |  |
|   |      | 910  |      |      |        |             |      | _   | 30<br> |      |       |       |              |      | 95     | -   |     |      |     |  |  |
| GAAT                                    |      |      |      |      |        |             |      |     |        |      |       |       |              |      |        |     |     |      |     |  |  |
| Ι                                       | E    | N    | A    | T    | Q      | A           | Q    |     |        | С    | L     | L     | V            |      |        |     | G   | G    | A   |  |  |
|   |      | 970  |      |      |        |             |      | -   | 90     |      |       |       |              |      | 101    |     |     |      |     |  |  |
| CTGC                                    | GGA  | CTG  | CCT  | GGC  | GGA    | GAC         | CCT  | GGA | AGA    | CAC  | TCT   | GGC   | CCC          | AGG  | GAG    | TGG | GGG | AGC  | CA  |  |  |
| A                                       | D    | С    | L    | A    | E      | T           | L    | E   | D      | T    | L     | A     | P            | G    | S      | G   | G   | A    | R   |  |  |
|   | 1    | .030 |      |      |        |             |      | 10  | 50     |      |       |       |              |      | 107    | 0   |     |      |     |  |  |
| GGCA                                    | AGG  | CGA  | 4GC  | CCG  | AGA:   | rcg         | TAA  | CAG | GCG    | TTT  | CTT   | TCC   | CAA          | AGG  | GGA    | CCT | TGA | GGT  | CC  |  |  |

Fig. 9 / continue n 1

|   |   |  | _  |   |  | _  |   | _   | _  | _  |  | _   |  |  | -   | .,  |   |
|---|---|--|--|---|--|--|---|---|--|--|--|---|--|--|---|---|---|
| . Q   | G E   | A  | R  | D I   | R I  |  | R R   | F   | F  | P  | K  | G   | D 1  | b  | Ε   | V   | Ļ   |
| moen.   | 1090  | -  |  |   |  |  | 1110  |   |  |  |  |   | 1130   | n n er   | mC=   | nar~  | mc  |
|   | GGCCCA  |  |  |   |  |  |   |   |  |  |  |   |  |  | s<br>S  | S   | E   |
| Q   | A Q   | ٧  | E  | R   | IM   |  |   | K   | E  | L  | Г  | T .   | -  | ľ  | 5   | 5   | E,  |
|   | 1150  |  |  | <b></b>   | ~~ ~-  |  | 1170  | nmm.  |  |  | -  |   | 1190   | ~~~  | mer.  | ncc   | ~~  |
|   | TGGGTC  |  |  |   |  |  |   |   |  |  |  |   |  |  | C   |   |   |
| D   | G S   | E  | E  | F   | E I  | _  | I V   | ъ   | K  | A  | ע  |   | K 3  | H  | C   | G   | S   |
|   | 1210  |  |  |   |  |  | 1230  |   |  |  |  |   | 1250   |  |   | ~~  |   |
|   | GGAGGC  |  |  |   |  | _  |   |   |  |  |  |   |  |  |   |   |   |
| S   | E A   | S  | A  | Y   | LI   | )  | E L   | R   | L  | A  | ν  |   |  | N  | R   | V   | D   |
|   | 1270  |  |  |   |  |  | 1290  |   |  |  |  |   | 1310   |  |   |   |   |
|   | TGCCCA  |  |  |   |  |  |   |   |  |  |  |   |  |  |   |   |   |
| I   | A Q   | S  | E  | L   | F  | 3  | G D   | I   | Q  | W  | R  | -   | _  | H  | L   | E   | A   |
|   | 1330  |  |  |   |  |  | 1350  |   |  |  |  |   | 1370   |  |   |   |   |
|   | CCTCAT  |  |  |   |  |  |   |   |  |  |  |   |  |  |   |   |   |
| S   | L M   | D  | A  | L   | L  | Ŋ  | D R   | P   | E  | F  | ٧  | -   |  | L  | 1   | S   | H   |
|   | 1390  |  |  |   |  |  | 1410  | ~   |  |  |  |   | 1430   |  |   | ~~~   |   |
|   | CCTCAG  |  |  |   | -  |  |   |   |  |  |  |   |  | •  |   |   |   |
| G   | L S   | L  | G  | H   | F 1  | L  | T P   | M   | R  | L  | A  |   | L  | 1  | 5   | A   | A   |
| ~~~   | 1450  |  | om.  | n ma  |  |  | 1470  |   | ~~~  | -00  | -m-c-  |   | 1490   | CO7  |   | ח תר  | C N   |
| P   | CTCCAA<br>S N   | S  |  |   | R I  |  | LL  |   |  |  | S  | H<br>H  |  | G<br>A   | G   | T   | K   |
| -   | 1510  | 5  | L  | -   | K I  | N  | 1530  | ט   | Q  | A  | 3  |   | 1550   | •  | G   | •   | v   |
| nncc  | CCCAGC  | لاطماحات   | מממו   | ccc   | CCDC   | امت  |   | ر ماسات   | ~~~  | 2000   | ٠٠٠٠   |   |  | cco  | ים  | ucu   | cc  |
| A   | P A   | L  |  |   |  |  | A E   |   | R  |  |  |   | v  |  |   | Δ.  |   |
| •   | 1570  |  | K  | ۳.  | <b>G</b> ,   | •  | 1590  | ם   |  | -  | -  | _   | 1610   |  | 11  | ٠   | -   |
| TCAC  | GATGCT  | cene   | ccc  | מממי  | יבעייתי  | יבעד   |   | CD C  | ומיזיבי  | ~~~  | <u>ጉ</u> ሞር (  |   |  | GC(  | TG  | GGA   | cc.   |
| R   | M L   | L  |  |   | M (  |  | A P   | R   | Y  | P  | s  | G   |  | _  | W   | Δ.  | P   |
| •   | 1630  | _  | •  |   |  | _  | 1650  | **  | •  | •  | _  | _   | 1670   |  |   | _   | •   |
| CTCA  | CCCAGG  | CCAG   | GGC  | TTC   | ragg(  | GAG  |   | GTA:  | rcro   | CTC  | CTCC   |   |  |  | CAC   | CTC   | GC  |
| H   | P G   | Q  |  |   | G I  |  | SM  | Y   |  | L  | s  | D   |  | A  | T   | s   | P   |
|   | 1690  | _  |  |   | •  |  |   |   |  |  |  |   | 1730   |  |   |   |   |
|   |   |  |  |   |  |  | 1710  |   |  |  |  |   | T120   |  |   |   |   |
| CGCT  | CTCGCT  | GGA1   | rgci   | 'GGC  | cro  | SGG  |   | CCC   | CTG  | GAG  | CGA  |   |  |  | r <b>r</b> G  | GGC   | AC  |
| CGCT  | CTCGCT  | GGAT<br>D  |  |   | CTC  |  |   |   | CTG(   |  |  |   | GCTT   |  | rtg<br>W  | GGC<br>A  | AC<br>L   |
|   |   |  |  |   |  |  | CAGGC   |   |  |  |  | CT<br>L   | GCTT   | CT'I<br>L  |   |   |   |
| L   | S L   | D  | A  | G   | L (  | G  | CAGGC<br>Q A<br>1770  | P   | W  | s  | D  | L   | GCTT<br>L<br>1790  | CT'I<br>L  | W   | A   | L   |
| L   | S L<br>1750   | D  | a<br>GCA   | CAG<br>G  | l (  | G<br>GCC   | CAGGC<br>Q A<br>1770  | P   | W<br>CTG   | s  | D  | L   | GCTT<br>L<br>1790<br>TTCC  | CT'I<br>L  | W   | A   | L<br>TT   |
| l<br>TGTT   | s l<br>1750<br>GCTGAA   | D<br>CAGO<br>R   | a<br>GCA   | CAG<br>G  | l (  | G<br>GCC   | CAGGC<br>Q A<br>1770<br>CATGTA  | P<br>CTT  | W<br>CTG   | s<br>Gga   | D<br>Sat(  | L<br>G<br>G   | GCTT<br>L<br>1790<br>TTCC  | CTT<br>L<br>AAT  | W<br>IGC  | a<br>agt  | L<br>TT   |
| l<br>TGTT<br>L  | S L<br>1750<br>GCTGAA<br>L N  | D<br>CAGO<br>R   | A<br>EGCA<br>A   | G<br>ACAG<br>Q  | L (<br>ATG(<br>M )   | G<br>GCC<br>A  | CAGGC<br>Q A<br>1770<br>CATGTA<br>M Y<br>1830   | P<br>CTT<br>F   | W<br>CTG(<br>W   | S<br>GGA<br>E  | D<br>SATO<br>M   | L<br>G<br>G   | GCTT<br>L<br>1790<br>TTCC<br>S<br>1850   | CT'I<br>L<br>AAI   | W<br>IGC.<br>A  | a<br>agt<br>v                                   | L<br>TT<br>S  |
| l<br>TGTT<br>L  | S L<br>1750<br>GCTGAA<br>L N<br>1810  | D<br>CAGO<br>R   | A<br>EGCA<br>A   | g<br>LCAG<br>Q<br>TGT   | L (<br>ATG(<br>M /   | G<br>GCC<br>A  | CAGGC<br>Q A<br>1770<br>CATGTA<br>M Y<br>1830   | P<br>CTT<br>F   | W<br>CTG(<br>W<br>GAT(   | S<br>GGA<br>E  | D<br>SATO<br>M   | L<br>G<br>G   | GCTT<br>L<br>1790<br>TTCC<br>S<br>1850   | CTT<br>L<br>AAT<br>N   | W<br>IGC.<br>A  | a<br>agt<br>v                                   | L<br>TT<br>S  |
| TGTT L  | S L<br>1750<br>GCTGAA<br>L N<br>1810<br>AGCTCT  | D<br>CAGO<br>R<br>TGGO                                 | A<br>GCA<br>A<br>GCC   | g<br>LCAG<br>Q<br>TGT   | L (<br>ATG(<br>M )   | G<br>GCC<br>A<br>CTC   | CAGGC<br>Q A<br>1770<br>CATGTA<br>M Y<br>1830<br>CCTCCG   | P<br>CTT(<br>F<br>GGT(                                | W<br>CTG(<br>W<br>GAT(   | S<br>GGA<br>E<br>GGC                                     | D<br>GATO<br>M<br>ACGO   | CT<br>G<br>CT<br>L  | I<br>1790<br>TTCC<br>S<br>1850   | CTT<br>L<br>PAAT<br>N<br>CCT   | W<br>IGC<br>A<br>IGA  | a<br>agt<br>v<br>ccc                            | L<br>TT<br>S<br>TG  |
| L<br>TGTT<br>L<br>CCTC                                      | S L<br>1750<br>GCTGAA<br>L N<br>1810<br>AGCTCT<br>A L   | D<br>CAGO<br>R<br>TGGO<br>G                            | A<br>GCA<br>A<br>GGCC<br>A   | G<br>ACAG<br>Q<br>TGT<br>C  | L ( ATGO M ) TTGO L ]  | G<br>GCC<br>A<br>CTC<br>L  | CAGGC<br>Q A<br>1770<br>CATGTA<br>M Y<br>1830<br>CCTCCG<br>L R<br>1890  | P<br>CTT<br>F<br>GGT<br>V                             | W<br>CTGG<br>W<br>GATG<br>M  | S<br>E<br>E<br>E<br>GGCI<br>A                            | D<br>SATO<br>M<br>ACGO<br>R  | CT<br>G<br>G<br>CT<br>L   | L<br>1790<br>TTCC<br>S<br>1850<br>GGAG<br>E  | CTT<br>L<br>AAT<br>N<br>CCT  | W<br>PGC.<br>A<br>PGA:<br>D   | A<br>V<br>CGC<br>A                              | L<br>TT<br>S<br>TG<br>E   |
| L<br>TGTT<br>L<br>CCTC                                      | S L<br>1750<br>GCTGAA<br>L N<br>1810<br>AGCTCT<br>A L<br>1870   | D<br>CAGO<br>R<br>TGGO<br>G                            | A<br>GCA<br>A<br>GGCC<br>A   | G<br>ACAG<br>Q<br>CTGT<br>C   | L ( ATGO M ) TTGO L )  | G<br>GCC<br>A<br>CTC<br>L  | CAGGC<br>Q A<br>1770<br>CATGTA<br>M Y<br>1830<br>CCTCCG<br>L R<br>1890  | P<br>CTT<br>F<br>GGT<br>V                             | W<br>CTGG<br>W<br>GATG<br>M  | S<br>E<br>E<br>E<br>GGCI<br>A                            | D<br>SATO<br>M<br>ACGO<br>R  | L<br>EGG<br>G<br>CCT<br>L<br>SAT<br>M   | TGCTT L 1790 TTCC S 1850 GGAG E 1910 GGGC G  | CTTL<br>L<br>AAT<br>N<br>CCTL<br>P   | W<br>PGC.<br>A<br>PGA:<br>D   | A<br>V<br>CGC<br>A                              | L<br>TT<br>S<br>TG<br>E   |
| TGTT L CCTC S   | S L<br>1750<br>GCTGAA<br>L N<br>1810<br>AGCTCT<br>A L<br>1870   | D<br>CAGO<br>R<br>TGGO<br>G<br>ACGO                    | A<br>GGCA<br>A<br>GGCC<br>A  | g<br>ACAG<br>Q<br>TGT<br>C  | L ( ATGO M ) TTGO L )  | G<br>GCC<br>A<br>CTC<br>L  | SCAGGC Q A 1770 LATGTA M Y 1830 SCTCCG L R 1890 SGCGTT  | P<br>CTT<br>F<br>GGT<br>V<br>CAA                      | W<br>CTGC<br>W<br>GATC<br>M  | S<br>E<br>E<br>E<br>GGC!<br>A<br>TGA(                    | D<br>EATO<br>M<br>ACGO<br>R<br>EGGG  | L<br>EGG<br>G<br>CCT<br>L<br>SAT<br>M   | TTCC<br>S<br>1850<br>GGAG<br>E<br>1910   | CTTL<br>L<br>AAT<br>N<br>CCTL<br>P   | W<br>PGC<br>A<br>PGA<br>D   | A<br>V<br>CGC<br>A                              | TT S TG E   |
| TGTT L CCTC S AGGA  | S L<br>1750<br>GCTGAA<br>L N<br>1810<br>AGCTCT<br>A L<br>1870<br>AGCAGC<br>A A<br>1930  | D<br>CAGO<br>R<br>TGGO<br>G<br>ACGO<br>R               | A<br>GGCA<br>A<br>GGCC<br>A<br>GAGG<br>R   | G<br>ACAG<br>Q<br>ETGT<br>C<br>C<br>BAAA<br>K   | L ( ATGO M A TTGO L ) AGACO D )  | G<br>GCCCAA<br>CTCC<br>L<br>CTCC   | ECAGGC Q A 1770 CATGTA M Y 1830 SCTCCG L R 1890 SEGCGTT A F 1950 SEGTGAG  | P<br>CTT(<br>F<br>GGT(<br>V<br>CAA)<br>K              | W<br>CTGC<br>W<br>GATC<br>M<br>GTT:  | S<br>EGAG<br>E<br>GGC!<br>A<br>TGAG<br>E                 | D<br>SATO<br>M<br>ACGO<br>R<br>SGGG<br>G   | CCT L CCT L GAT M   | GCTT L 1790 TTCC S 1850 GGAG E 1910 GGGC G 1970  | CTTL AAAT N CCCTL P CGTTL CGTT | W<br>IGC.<br>A<br>IGA<br>D<br>D   | A AGT V CGC A CCT L                             | TT S TG E CT F  |
| TGTT L CCTC S AGGA  | S L<br>1750<br>GCTGAA<br>L N<br>1810<br>AGCTCT<br>A L<br>1870<br>GGCAGC<br>A A<br>1930<br>CGGAGTG   | D<br>CAGG<br>R<br>TGGG<br>G<br>ACGG<br>R<br>CTAT       | A<br>GGCA<br>A<br>GGCC<br>A<br>GAGG<br>R   | G<br>ACAG<br>Q<br>ETGT<br>C<br>C<br>BAAA<br>K   | L ( ATGO M A TTGO L ) AGACO D )  | G<br>GCCCAA<br>CTCC<br>L<br>CTCC   | ECAGGC Q A 1770 CATGTA M Y 1830 SCTCCG L R 1890 SGCGTT A F 1950 SGTGAG V R  | P<br>CTT(<br>F<br>GGT(<br>V<br>CAA)<br>K              | W<br>CTGC<br>W<br>GATC<br>M<br>GTT:  | S<br>EGAG<br>E<br>GGC!<br>A<br>TGAG<br>E                 | D<br>SATO<br>M<br>ACGO<br>R<br>SGGG<br>G   | L<br>EGG<br>G<br>CCT<br>L<br>GAT<br>M   | IGCTT L 1790 TTCC S 1850 GGAG E 1910 GGGC G 1970 CCTC L  | CTTL LAATIN CCTL P CGTTL CGTTL R   | W<br>IGC.<br>A<br>IGA<br>D<br>D   | A AGT V CGC A CCT L                             | TT S TG E CT F  |
| TGTT L CCTC S AGGA E TTGG                                   | S L<br>1750<br>GCTGAA<br>L N<br>1810<br>AGCTCT<br>A L<br>1870<br>GGCAGC<br>A A<br>1930<br>CGAGTG<br>E C                                   | D<br>CAGO<br>R<br>TGGO<br>G<br>ACGO<br>R<br>CTAT       | A<br>A<br>SGCCO<br>A<br>BAGGO<br>R   | G<br>ACAG<br>Q<br>ETGT<br>C<br>C<br>EAAAA<br>K  | L ( EATGO M A TTIGO L 1 LGACO D 1 EAGTO S I  | G<br>GCCCAA<br>CTCC<br>L<br>CTCC<br>GACC   | SCAGGC Q A 1770 CATGTA M Y 1830 SCTCCG L R 1890 SGCGTT A F 1950 SGTGAG V R 2010   | P<br>CTT'<br>F<br>GGT'<br>CAAA<br>K<br>GGCC           | W CTGC W GATC M GTT F TGCC   | S<br>GGAG<br>E<br>GGCI<br>A<br>TGAG<br>E                 | D  GATO  M  ACGO  R  GGG  GCTO  L  | ECT<br>L<br>SAT<br>M<br>CCT<br>L  | GCTT L 1790 TTCC S 1850 GGAG E 1910 GGGC G 1970 CCTC L   | CTTL<br>L<br>AAT<br>N<br>CCTT<br>V<br>CGTT   | W<br>IGC.<br>A<br>IGA<br>D<br>IGA<br>ICG<br>R   | A AGT V CGC A CCTG C                            | TT S TG E CT F  |
| TGTT L CCTC S AGGA E TTGG G                                 | S L 1750 GCTGAA L N 1810 AGCTCT A L 1870 AGCAGC A A 1930 CCGAGTG E C 1990 CTGGGGG   | D CAGGOR R TGGGG R ACGGOR Y Y GGAT                     | A GGGGG A A GGGGGG R R CGGGG   | G<br>ACAG<br>Q<br>TTGT<br>C<br>C<br>K<br>K<br>K<br>AGC<br>S   | L () TTTGGACGACGACGACGACGACGACGACGACGACGACGACGAC   | G<br>GCC<br>A<br>CTC<br>L<br>CTC<br>E<br>A<br>GAG<br>E   | SCAGGC Q A 1770 CATGTA M Y 1830 SCTCCG L R 1890 SGCGTT A F 1950 SGTGAG V R 2010 CCAGCT  | P<br>CTTC<br>F<br>GGTC<br>V<br>CAAA<br>K<br>GGCC      | W CTGC W GATC M GTT F TGCC A CATC  | S<br>GGAG<br>E<br>GGGG<br>R<br>GGGA                      | D GATO M ACGO R GGGGG G G AGCTO L  | CCT<br>L<br>SAT<br>M<br>CCT<br>L  | E 1910<br>EGGAG<br>E 1910<br>EGGAG<br>E 1910<br>EGGCC<br>L 2030  | CTTL LAATIN CCCTL P GGTTL R CCGTL  | W<br>IGC<br>A<br>IGA<br>D<br>IGA<br>ICG<br>R  | A AGT V CGC A CTT                               | TT S TG E CT F  |
| TGTT L CCTC S AGGA E TTGG G                                 | S L<br>1750<br>GCTGAA<br>L N<br>1810<br>AGCTCT<br>A L<br>1870<br>GGCAGC<br>A A<br>1930<br>CGAGTG<br>E C                                   | D CAGGOR R TGGGG R ACGGOR Y Y GGAT                     | A GGGGG A A GGGGGG R R CGGGG   | G<br>ACAG<br>Q<br>TTGT<br>C<br>C<br>K<br>K<br>K<br>AGC<br>S   | L () TTTGGACGACGACGACGACGACGACGACGACGACGACGACGAC   | G<br>GCC<br>A<br>CTC<br>L<br>CTC<br>E<br>A<br>GAG<br>E   | SCAGGC Q A 1770 CATGTA M Y 1830 SCTCCG L R 1890 SGCGTT A F 1950 SGTGAG V R 2010 CCAGCT  | P<br>CTTC<br>F<br>GGTC<br>V<br>CAAA<br>K<br>GGCC      | W CTGC W GATC M GTT F TGCC A CATC  | S<br>GGAG<br>E<br>GGGG<br>R<br>GGGA                      | D GATO M ACGO R GGGGG G G AGCTO L  | CCT<br>L<br>SAT<br>M<br>CCT<br>L  | E 1910<br>EGGAG<br>E 1910<br>EGGAG<br>E 1910<br>EGGCC<br>L 2030  | CTTL LAATIN CCCTL P GGTTL R CCGTL  | W<br>IGC<br>A<br>IGA<br>D<br>IGA<br>ICG<br>R  | A AGT V CGC A CTT                               | TT S TG E CT F  |
| L TGTT L CCTC S AGGA E TTGG G CGCT L                        | S L<br>1750<br>GCTGAA<br>L N<br>1810<br>AGCTCT<br>A L<br>1870<br>AGCAGC<br>A A<br>1930<br>CCGAGTG<br>E C<br>1990<br>CTGGGG<br>W G<br>2050 | D CAGO R G G ACGO R CTAT Y                             | A GGGCC A AAAGGC R CGGCC A   | G<br>ACAG<br>Q<br>CTGT<br>C<br>BAAA<br>K<br>K<br>CAGC<br>S<br>CACT  | L ()  ATGG M )  TTGG GACCO D I  AGTC S I  TGCC C I   | G<br>GCC<br>A<br>CTC<br>L<br>CTC<br>L  | SCAGGC Q A 1770 CATGTA M Y 1830 SCTCCG L R 1890 SGCGTT A F 1950 SGTGAG V R 2010 CCAGCT Q L 2070   | P<br>CTTC<br>F<br>GGTC<br>V<br>CAAA<br>K<br>GGCC<br>A | W CTGC W GATC M F TGCC A CATC  | S GGAG E GGCI A TGAG E CCGG R .                          | D GATO M ACGO R GGGG G CCTO L AGCO A   | CCT<br>L<br>SGG<br>G<br>CCT<br>L<br>SAT<br>M<br>CCT<br>L                                  | GCTT L 1790 TTCC S 1850 GGGAG E 1910 GGGC G 1970 CCCC L 2030 CCGCC A 2090                                      | CTT<br>L<br>AAAT<br>N<br>CCTT<br>P<br>CGTT<br>R  | W<br>PGC<br>A<br>PGA<br>D<br>PCG<br>R   | A AGT V CGC A CCT CTG C CTF                     | TT S TG E CT F  |
| L TGTT L CCTC S AGGA E TTGG G CGCT L                        | S L 1750 GCTGAA L N 1810 AGCTCT A L 1870 AGCAGC CA A 1930 CCAGGTG E C 1990 CCTGGGG W G 2050 CCCAGGA                                       | D CAGO R TGGG R CTAT Y GGGAT D TGGGG                   | A GGCA A GGCC A A GGCC  | G ACAG Q CTGT C C CAGC S CACT T   | L ()  ATGGACO  B I  AGTTGCC  C I   | G<br>GCC<br>A<br>CTC<br>L<br>CTC<br>L  | SCAGGC Q A 1770 CATGTA M Y 1830 SCTCCG L R 1890 SGCGTT A F 1950 SGTGAG V R 2010 CCAGCT Q L 2070 SCTGAC  | P<br>CTTC<br>F<br>GGTC<br>K<br>CAAN<br>K<br>GGCC<br>A | W CTGC W GATC M GTT F CATC A CATC  | S GGA E GGCI A TGA E CCGG R . GCCAI                      | D  SATO M  M  ACGO R  CCTO L  AGCO A  GTGO   | CT<br>L<br>EGG<br>G<br>CT<br>L<br>EAT<br>M<br>CT<br>L                                     | GCTT L 1790 TTCC S 1850 GGGAG E 1910 GGGCC G 2030 CCCC A 2090 AGAT   | CTTL AAAT N CCCTL P CCGTL R CCGTL R  | W<br>PIGC.<br>A<br>PIGA.<br>D<br>PICG.<br>R<br>PIGC.<br>A   | A AGT V CCCC A CCT C CTC C CTT F                | TT S TG E CT F CCC P  |
| L TGTT L CCTC S AGGA E TTGG G CGCT L                        | S L 1750 GCTGAA L N 1810 AGCTCT A L 1870 AGCAGC CCAGGTG E C 1990 CCTGGGG W G 2050 CCAGGAGC  | D CAGO R TGGG R CTAT Y GGGAT D TGGGG                   | A GGCA A GGCC A A GGCC  | G ACAG Q CTGT C C CAGC S CACT T   | L ()  ATGGACO  B I  AGTTGCC  C I   | G<br>GCC<br>A<br>CTC<br>L<br>CTC<br>L  | SCAGGC Q A 1770 CATGTA M Y 1830 SCTCCG L R 1890 SGCGTT A F 1950 SGTGAG V R 2010 CCAGCT Q L 2070 SCTGAC  | P<br>CTTC<br>F<br>GGTC<br>K<br>CAAN<br>K<br>GGCC<br>A | W CTGC W GATC M GTT F CATC A CATC  | S GGA E GGCI A TGA E CCGG R . GCCAI                      | D  SATO M  M  ACGO R  CCTO L  AGCO A  GTGO   | CT<br>L<br>EGG<br>G<br>CT<br>L<br>EAT<br>M<br>CT<br>L                                     | GCTT L 1790 TTCC S 1850 GGGAG E 1910 GGGCC G 2030 CCCC A 2090 AGAT   | CTTL AAAT N CCCTL P CCGTL R CCGTL R  | W<br>PIGC.<br>A<br>PIGA.<br>D<br>PICG.<br>R<br>PIGC.<br>A   | A AGT V CCCC A CCT C CTC C CTT F                | TT S TG E CT F CCC P  |
| TGTT L CCTC S AGGA E TTGG G CGCT L                          | S L 1750 GCTGAA L N 1810 GGCTGTA A L 1870 GGCAGC A A 1930 GCGAGTG E C 1990 CCTGGGG W G 2050 CCCAGGA Q D 2110                              | D CAGO   | A GGCA A GGCCA R R GGCCA R CGCCA V   | G<br>ACAG<br>Q<br>CTGT<br>C<br>C<br>EAAA<br>K<br>K<br>AGC<br>S<br>LACT<br>T   | L ( ATGO M ) TTTGG L 1 GACO D 1 TAGTO C I TTGCC S 1  | G<br>GCCAA<br>CTCC<br>L<br>CTCC<br>L   | SCAGGC Q A 1770 CATGTA M Y 1830 SCTCCG L R 1890 SGCGTT A F 1950 SGCGAG V R 2010 CCAGCT Q L 2070 SCTGAC L T 2130   | P<br>CTTC<br>F<br>GGTC<br>V<br>CAAN<br>K<br>GGCC<br>A | W CTGC W GATC M GTT F TGCC A CATC M GAAC K   | S GGA E GGC A TGA C CCGC R . GCCA C GCTG W               | D  SATO M  ACGO R  CCTO L  AGCO A  G  W  | ECT<br>L<br>SAT<br>M<br>CCT<br>L<br>FGA<br>D  | GCTT L 1790 TTCC S 1850 GGAG E 1910 GGGC G L 2030 CCCC A 2090 AGAT D 2150                                      | CTTL  AAATA  N  CCTL  P  CCTL  R  CCGTL  R  ATCL  M  | W<br>IGC<br>A<br>IGA<br>D<br>ICG<br>R<br>IGC<br>A   | A AGT V CCC A CCT C CTG C CTG C CAG             | TT S TG E CT F CC P CT F CA T   |
| TGTT L CCTC S AGGA E TTGG G CGCT L                          | S L 1750 GCTGAA L N 1810 GGCTGTA A L 1870 GGCAGC A A 1930 GCGAGTG E C 1990 CCTGGGG W G 2050 CCCAGGA Q D 2110 ACCCAT                       | D CAGGO R TTGGG G R ACGGO R CTAT Y GGGAT G G G G CTGGG | A GGCA A GGCCA R GGCCA R GGCCA C C GGCCA C GGCCCA C GGCCC | G ACAG Q CTGT C C K K CAGC S CACT T ACAG  | L ( )  L ( )  GAC( )  C I  TECCC C I  S I  GGTTC C S I   | G<br>GCCAA<br>CTC<br>L<br>CTC<br>CTC   | SCAGGC Q A 1770 CATGTA M Y 1830 SCTCCG L R 1890 SGCGTT A F 1950 SGCGAG V R 2010 CCAGCT Q L 2070 SCTGAC L T 2130   | P CTTC F GGTC V CAAN K GGCC A ACAC Q CTTC             | W CTGG W GATC M GTT TGGG A CATC M GAAG K TTGG  | S GGA E GGCI A TGA E CCGC R . GGCAI GGCAI W CCCC         | D  SATO M  ACGO R  SEGGO G  CCTO L  AGCO W  FCCO   | CCT<br>L<br>SGG<br>G<br>CCT<br>L<br>SAT<br>M<br>CCT<br>L<br>CGA<br>D<br>GGG<br>G          | GCTT L 1790 TTCC S 1850 GGAG E 1910 GGGC C L 2030 CCCC A 2090 AGAT D 2150 CCATC                                | CTTL AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA   | W IGC A IGA IGA ICG R IGC A IGC A IGC A   | A AGT V CCCC A CCTT F CAG S CCCC                | L TT S TG E CT F CC P CT F CC T T CC C  |
| TGTT L CCTC S AGGA E TTGG G CGCT L                          | S L 1750 GCTGAA L N 1810 GGCTGTA A L 1870 GGCAGC A A 1930 GCGAGTG E C 1990 CCTGGGG W G 2050 CCCAGGA Q D 2110                              | D CAGGO R TTGGG G R ACGGO R CTAT Y GGGAT G G G G CTGGG | A GGCA A GGCCA R GGCCA R GGCCA C C GGCCA C GGCCCA C GGCCC | G ACAG Q CTGT C C K K CAGC S CACT T ACAG  | L ( )  L ( )  GAC( )  C I  TECCC C I  S I  GGTTC C S I   | G<br>GCCAA<br>CTC<br>L<br>CTC<br>CTC   | SCAGGC Q A 1770 CATGTA M Y 1830 SCTCCG L R 1890 SGCGTT A F 1950 SGTGAG V R 2010 CCAGCT Q L 2070 SCTGAC L T 2130 CGCCTT A F                              | P CTTC F GGTC V CAAN K GGCC A ACAC Q CTTC             | W CTGG W GATC M GTT IGCO A CATC M GAAG K ITGG  | S GGA E GGCI A TGA E CCGC R . GGCAI GGCAI W CCCC         | D  SATO M  ACGO R  SEGGO G  CCTO L  AGCO W  FCCO   | CCT<br>L<br>SAT<br>M<br>CCT<br>L<br>FGA<br>D<br>GGG<br>G                                  | GCTT L 1790 TTCC S 1850 GGAG E 1910 GGGC L 2030 CCCC A 2090 AAGAT D 2150 CCATC I                               | CTTL AAAT N CCCTL P CCGTL R CCGTL R CCGTL R TAAAT  | W IGC A IGA IGA ICG R IGC A IGC A IGC A   | A AGT V CCCC A CCTT F CAG S CCCC                | L TT S TG E CT F CC P CT F CC T T CC C  |
| TGTT L CCTC S AGGA E TTGG G CGCT L TTGC A                   | S L 1750 GCTGAA L N 1810 GGCTGTA A L 1870 GGCAGC A A 1930 GCGAGTG E C 1990 CCTGGGG W G 2050 CCCAGGA Q D 2110 ACCCATT P I 2170             | D CAGO R G G G G G G G G G G G G G G G G G G           | A GGGCC A A AAGGC R CGGCC A V CGGCCC A   | G ACAGO Q CTGT C C CAGCO S LACT T ACAG  | L (  ATGGACC  B I  AGTC  C I  TTCC  C I  TTCC  V I   | G<br>GCCOA<br>L<br>CTCO<br>L<br>CTCO<br>L<br>CTCO  | SCAGGC Q A 1770 CATGTA M Y 1830 SCTCCG L R 1890 SGCGTT A F 1950 SGCGAG V R 2010 CCAGCT Q L 2070 SCTGAC L T 2130 CGCCTT A F                              | P CTTC F GGTC CAAC K GGCC A CGCC A CCTTC              | W CTGG W GATG M F F CATGG K CTGG C   | S GGGA E GGGCI A TGA E CCGG R . GCCAI Q GTGG W           | D  GATO  ACGO  R  CCTO  L  AGCO  A  CTCO  P  | ECT<br>L<br>SEGG<br>G<br>CCT<br>M<br>CCT<br>L<br>FGA<br>D<br>GGG<br>G                     | GCTT I 1790 TTCC S 1850 GGAG E 1910 GGGCC L 2030 CCCCC A 2090 AGAT D 2150 CCATC I 2210                         | CTTL AAAT N CCCTL P CCGTL R CCGTL R CCGTL R TAAAT  | W<br>PGC:<br>A<br>PGA:<br>D<br>PGG:<br>R<br>PGC:<br>A<br>PGC:<br>A  | A AGT V CGC A CTT F CAG S CCG R                 | TT S TG E CT F CC L   |
| TGTT L CCTC S AGGA E TTGG G CGCT L TTGC A                   | S L 1750 GCTGAA L N 1810 GGCTGTA A L 1870 GGCAGC A A 1930 CCGAGTG E C 1990 CCAGGA Q D 2110 ACCCATT P I 2170 CCACCTT                       | D CAGO   | A GGCC A GGCC A CGCC A CGCC A CGCC A GGCC A  | G ACAGO Q CTTGT C C EAAAA K CAGO S LACT T ACAG Q L LTCAG  | L ( )  ATGCC  B I  AGTC  C I  TTCC  TTCC | G<br>GCCAA<br>CTCC<br>L<br>CTCC<br>L<br>CTCC<br>L  | SCAGGC Q A 1770 CATGTA M Y 1830 SCTCCG L R 1890 SGCGTT A F 1950 SGCGAG V R 2010 CCAGCT Q L 2070 SCTGAC L T 2130 CGCCTT A F 2190 SGAGCC                  | P CTTC F GGTC V CAAA K GGCC A CGCC A CCTTC F CACAC    | W CTGC W GATC M GTT TGC A CATC M CATC CATC A CATC CATC CATC CA   | S GGGA E GGGA  TGA C C C C C C C C C C C C C C C C C C C | D  GACGO R  GGGGG G  GCCTO L  AGCO R  FCCI R   | ECT<br>L<br>SAT<br>M<br>CT<br>L<br>FGA<br>D<br>SGG<br>G                                   | GCTT I 1790 TTCC S 1850 GGAG E 1910 GGGCC L 2030 CCCCC A 2090 AGAT D 12210 AGAG                                | CTTL AAT N CCTL P GTTL CGTL R CGTL R CGTL TAC Y  | W IGC A IGA D IGA ICG R IGC A | A AGT V CCCC A CCTC C CTG C CAG S CCAG R CAT    | TT S TG E CC P CT F CC L CG L   |
| TGTT L CCTC S AGGA E TTGG G CGCT L TTGC A                   | S L 1750 GCTGAA L N 1810 GGCTGTA A L 1870 GGCAGC A A 1930 GCGAGTG E C 1990 CCTGGGG W G 2050 CCCAGGA Q D 2110 ACCCATT P I 2170             | D CAGO   | A GGCC A GGCC A CGCC A CGCC A CGCC A GGCC A  | G ACAGO Q CTTGT C C EAAAA K CAGO S LACT T ACAG Q L LTCAG  | L ( )  ATGCC  B I  AGTC  C I  TTCC  TTCC | G<br>GCCAA<br>CTCC<br>L<br>CTCC<br>L<br>CTCC<br>L  | SCAGGC Q A 1770 CATGTA M Y 1830 SCTCCG L R 1890 SGCGTT A F 1950 SGCGAG V R 2010 CCAGCT Q L 2070 SCTGAC L T 2130 CGCCTT A F 2190 SGAGCC                  | P CTTC F GGTC V CAAA K GGCC A CGCC A CCTTC F CACAC    | W CTGC W GATC M GTT TGC A CATC M CATC CATC A CATC CATC CATC CA   | S GGGA E GGGA  TGA C C C C C C C C C C C C C C C C C C C | D  GACGO R  GGGGG G  GCCTO L  AGCO R  FCCI R   | ECT<br>L<br>SAT<br>M<br>CT<br>L<br>FGA<br>D<br>SGG<br>G                                   | GCTT I 1790 TTCC S 1850 GGAG E 1910 GGGCC L 2030 CCCCC A 2090 AGAT D 12210 AGAG                                | CTTL AAT N CCTL P GTTL CGTL R CGTL R CGTL TAC Y  | W IGC A IGA D IGA ICG R IGC A | A AGT V CCCC A CCTC C CTG C CAG S CCAG R CAT    | TT S TG E CC P CT F CC L CG L   |
| TGTT L CCTC S AGGA E TTGG G CGCT L TTGC A CTAC T            | S L 1750 GCTGAA L N 1810 GGCTGTA A L 1870 GGCAGC A A 1930 CCGAGTG E C 1990 CCAGGA Q D 2110 ACCCAT P I 2170 CACCTT T F 2230                | D CAGGOR R CTATA G G G G G G G G G G G G G G G G G G   | A GGCCA A GGCC | G ACAGO Q CTGT C GAAAA K CAGO S ACT T ACAG Q CCTG L   | L ( ATGCACO B I AGTCAC C I TTCCC S I TTCCC V I GAAC GAAC E I   | G<br>GCCAA<br>CTCC<br>L<br>CTCC<br>L<br>CTCC<br>L  | ECAGGC Q A 1770 CATGTA M Y 1830 ECTCCG L R 1890 EGCGTT A F 1950 ECTGAG V R 2010 CCAGCT Q L 2070 ECTGAC L T 2130 CCCCTT A F 2190 EGGGCCT E P 2250        | P CTTC F GGTC CAAC CTT: CACC T                        | CATCA  CA | S GGA E GGC A TGA E CCG R . GCC W CCC P GGGA E           | D  SATO  M  ACGO  R  SEGGO  G  CCTO  L  AGCO  W  FCCA  P   | ECT L SAT D SGG G L CT L SCT L SCT L  | GCTT I 1790 TTCC S 1850 GGAG E 1910 GGGC G 1970 CCCTC L 2030 CCGCC A 2090 AGAT D 12210 AGAG E 2210 AGAG E 2270 | CTTL AATAN CCCTR CCGTR ATAC Y TTTT   | W GCC A GCC R GCC A GCC A CAC T GCA D   | A AGT V CGC A CCT CTG C CTG F CAG S CCG R CAT   | TT S TTG E CT F CC P T CC L CC D  |
| TGTT L CCTC S AGGA E TTGG G CGCT L TTGC A CTAC T            | S L 1750 GCTGAA L N 1810 GGCAGC A L 1970 GGCAGC E C 1990 CCAGGA Q D 2110 ACCCAT P I 2170 CACCTT T F 2230 TGTCAT                           | D CAGO R TAGO G G CTGG W CAGO R                        | A GGCCA A GGCCCA A GGCCCCA A GGCCCCCA A GGCCCCCCA A GGCCCCCCA A GGCCCCCCA A GGCCCCCCA A GGCCCCCCCC   | G ACAGO S ACAT T ACAGO L ATCAGO S ACAT T ACAGO S ACAT T ACAGO S ACAT T ACAGO S ACAT S | L ( )  ATGGACO  B I  AGTO  C I  TTCK  S I  TTCK  TTCK  S I  TTCK  TTCK  S I  TTCK  TTCK  S I  TTCK  TTCK | G GCCAA CTCCL CTCCCCCCCC | ECAGGC Q A 1770 CATGTA M Y 1830 ECTCCG L R 1890 EGCGTT A F 1950 ECTGAG V R 2010 ECAGCT Q L 2070 ECTGAC L T 2130 ECCCTT A F 2190 EGGGCCT E P 2250 ECTCGG | P CTTC F CAAA CAAA ACAA Q CTTC F CACAC T GGACA        | W CTGC W GATC M GTT F F GATC A CATC A CATC A CATC C C C C C C C C  | S GGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG                  | D  GATO  M  ACGO  R  GCTO  L  GCTO  W  FCCO  P  GGAO  E  CCCO  CCCCO  CCCO  CC | CCT<br>L SAT<br>M CCT<br>L CCT<br>L CCT<br>L CCT<br>L CCT<br>L CCT<br>L CCT<br>L CCT      | GCTT I 1790 TTCC S 1850 GGAG E 1910 GGGC G 1970 CCTC L 2030 CCGC A 2090 AGAT D 12210 AGAG E 2270 CCGAG         | CTTL AATA N CCCTL CCTL CCTL R CCGTL R ATA Y TTTL F AAA   | W IGC A IGA D IGA C IGC A IGC | A AGT V CGC A CCT CTG C CTG F CAG S CCG R CAT M | L TT S TE E CT F CC P TE CC L |
| TGTT L CCTC S AGGA E TTGG G CGCT L TTGC A CTAC T            | S L 1750 GCTGAA L N 1810 AGCTCT A L 1870 AGCAGC A A 1930 CCAGGG W G 2050 CCAGGA Q D 2110 ACCCATT P I 2170 CACCTT T F 2230 TGTCAT          | D CAGO R TAGO G G CTGG W CAGO R                        | A GGCCA A GGCCCA A GGCCCCA A GGCCCCCA A GGCCCCCCA A GGCCCCCCA A GGCCCCCCA A GGCCCCCCA A GGCCCCCCCC   | G ACAGO S ACAT T ACAGO L ATCAGO S ACAT T ACAGO S ACAT T ACAGO S ACAT T ACAGO S ACAT S | L ( )  ATGGACO  B I  AGTO  C I  TTCK  S I  TTCK  TTCK  S I  TTCK  TTCK  S I  TTCK  TTCK  S I  TTCK  TTCK | G GCCAA CTCCL CTCCCCCCCC | ECAGGC Q A 1770 CATGTA M Y 1830 ECTCCG L R 1890 EGCGTT A F 1950 ECTGAG V R 2010 ECAGCT Q L 2070 ECTGAC L T 2130 ECCCTT A F 2190 EGGGCCT E P 2250 ECTCGG | P CTTC F CAAA CAAA ACAA Q CTTC F CACAC T GGACA        | W CTGC W GATC M GTT F F GATC A CATC A CATC A CATC C C C C C C C C  | S GGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG                  | D  GATO  M  ACGO  R  GCTO  L  GCTO  W  FCCO  P  GGAO  E  CCCO  CCCCO  CCCO  CC | CCT<br>L SAT<br>M CCT<br>L CCT<br>L CCT<br>L CCT<br>L CCT<br>L CCT<br>L CCT<br>L CCT      | GCTT I 1790 TTCC S 1850 GGAG E 1910 GGGC G 1970 CCTC L 2030 CCGC A 2090 AGAT D 12210 AGAG E 2270 CCGAG         | CTTL AATA N CCCTL CCTL CCTL R CCGTL R ATA Y TTTL F AAA   | W IGC A IGA D IGA C IGC A IGC | A AGT V CGC A CCT CTG C CTG F CAG S CCG R CAT M | L TT S TE E CT F CC P TE CC L |
| TGTT L CCTCC S AGGA E TTGGG G CGCT L TTGCC A CTACC T TCAT I | S L 1750 GCTGAA L N 1810 GGCAGC A L 1970 GGCAGC E C 1990 CCAGGA Q D 2110 ACCCAT P I 2170 CACCTT T F 2230 TGTCAT                           | D CAGO R TANT N  | A GGCCA A GGCCCA A GGCCCCA A GGCCCCCA A GGCCCCCCA A GGCCCCCCA A GGCCCCCCA A GGCCCCCCA A GGCCCCCCCC   | G ACAGO S ACAT T ACAGO L ATCAGO S ACAT T ACAGO S ACAT T ACAGO S ACAT T ACAGO S ACAT S | L ( )  ATGGACO  B I  AGTO  C I  TTCK  S I  TTCK  TTCK  S I  TTCK  TTCK  S I  TTCK  TTCK  S I  TTCK  TTCK | G GCCAA CTCCL CTCCCCCCCC | ECAGGC Q A 1770 CATGTA M Y 1830 ECTCCG L R 1890 EGCGTT A F 1950 ECTGAG V R 2010 ECAGCT Q L 2070 ECTGAC L T 2130 ECCCTT A F 2190 EGGGCCT E P 2250 ECTCGG | P CTTC F CAAA CAAA ACAA Q CTTC F CACAC T GGACA        | W CTGC W GATC M GTT F F GATC A CATC A CATC A CATC C C C C C C C C  | S GGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG                  | D  GATO  M  ACGO  R  GCTO  L  GCTO  W  FCCO  P  GGAO  E  CCCO  CCCCO  CCCO  CC | CCT<br>L<br>SGG<br>G<br>CCT<br>L<br>SAT<br>M<br>CCT<br>L<br>GGG<br>G<br>G<br>L<br>CT<br>L | GCTT I 1790 TTCC S 1850 GGAG E 1910 GGGC G 1970 CCTC L 2030 CCGC A 2090 AGAT D 12210 AGAG E 2270 CCGAG         | CTTL AAAT N CCTL GTTL GTTL GTTL GTTL GTTL GTTL GTTL  | W IGC A IGA D IGA C IGC A IGC | A AGT V CGC A CCT CTG C CTG F CAG S CCG R CAT M | L TT S TE E CT F CC P TE CC L |

PCT/EP01/08309

Fig. 9 / continue in 2

| TEGG  |  |  |  |   |   |   |   |  |  |  |   |  |   |  |  |  |  |
|---|--|--|--|---|---|---|---|--|--|--|---|--|---|--|--|--|--|
|   | GGTCCC   | GCGC   | CAC  | STC   |   |   |   |  |  |  |   |  | •   |  |  |  |  |
| G   | A b  | R  | Q  | s   | G   | R   | P G   | С  | С  | G  | G   | R  | С   | G  | G  | R  | R  |
|   | 2350   |  |  |   |   |   | 2370  |  |  |  |   |  | 2390  |  |  |  |  |
| GGTG  | CCTACG   | CCGC   | TGC  | STTC  | CAC   | CTT   | TGGGG   | CGTY                                       | SCC  | GGT  | SAC   | TAC  | CTTC  | ATG  | GGC  | AA.  | CG   |
| С   | L R  | R  | W  | F   | H   | F   | W G   | ν  | P  | ν  | T   | I  | F   | M  | G  | N  | V  |
|   | 2410   |  |  |   |   |   | 2430  |  |  |  |   |  | 2450  | i  |  |  |  |
| TGGT  | CAGCTA   | CCTC   | CTO  | TTC   | CT  | 3CT(  | CTTTT   | CTC  | 3CG(   | GGT  | CT(   | GCT  | CGTG  | GAT  | TT   | CCA  | GC   |
| v   | S Y  | L  | L  |   |   | L   | L F   | S  |  |  | L   | L  |   | D  | F  |  | P  |
| •   | 2470   |  | .,   | •   | -   | -   | 2490  | •  | ••   | •  | ~   | -  | 2510  | _  | _  | ~  |  |
| ~~~   | GCCGCC   | 000  | •  |   | ~~~   | ~~m   |   | COD N.                                     | nam  | ~~~  | ~~~   | win  |   |  | : Crite  | -mc  | ~  |
|   |  |  |  |   |   |   |   |  |  |  |   |  |   |  |  |  |  |
| A   | P P  | G  | S  | L   | E   | L   |   | Y  | F  | W  | A   | F  | -   | _  | L  | С  | E  |
|   | 2530   |  |  |   |   |   | 2550  |  |  |  |   |  | 2570  |  |  |  |  |
| AGGA  | ACTGCG   | CCA  | GGG  |   | -   |   |   |  |  |  |   |  |   |  |  |  |  |
| E   | L R  | Q  | G  | L   | S   | G   | G G   | G  | S  | L  | A   | S  | G   | G  | P  | G  | P  |
|   | 2590   |  |  |   |   |   | 2610  |  |  |  |   |  | 2630  | )  |  |  |  |
| CTGG  | CCATGC   | CTC  | ACT  | GAG   | CCA   | GCG   | CTGCG   | CCT  | CTA  | CCT  | CGC   | CGF  | CAGO  | CTG  | SAA  | CCA  | GT.  |
| G   | н а  | s  | L  | 3   | Q   | R   | L R   | L  | Y  | L  | A   | D  | S   | W  | N  | Q  | С  |
|   | 2650   |  |  |   |   |   | 2670  |  |  |  |   |  | 2690  | )  |  |  |  |
| GCGA  | CCTAGT   | GGC  | rcr  | CAC   | CTG   | CTT   | CTCCI   | GGG  | CGT  | GGG  | CTG   | cce  | GCT   | SAC  | CCC  | GGG  | TT   |
| D   | L V  |  |  | T   | С   |   | L L   | G  |  |  |   |  | L   | T  |  | G  |  |
| _   | 2710   |  | _  | •   | •   | _   | 2730  | _  | •  | •  | •   |  | 2750  | _  | _  | -  | _  |
| mcm s   | CCACCT   |  | ~~~  | ~ N ~!                                      | mem   | C THE   |   | CCN  | C-THIT   | CNE  | CCIP  | The Party  |   |  | 200  | C-T  | ec.  |
|   |  |  |  |   | V   | L<br>L  | CI  | D  | F  |  | v   |  | T   |  | R  | L  | L  |
| ¥   | H L  | G  | R  | T   | v   | ъ   |   | ע  | E  | £1   | ٧   | -  | _   | •  | ~  |  |  |
| _   | 2770   |  |  |   |   |   | 2790  |  |  |  |   |  | 2810  | _  |  |  |  |
| TTCA  | CATCTT   |  |  |   |   |   |   |  |  |  |   |  |   |  |  |  |  |
| H   | I F  | T  | V  | И   | K   | Q   |   | P  | K  | I  | v   | I  | V   |  | K  | M  | M  |
|   | 2830   |  |  |   |   |   | 2850  |  |  |  |   |  | 287   |  |  |  |  |
| TGAA  | LGGACGT  | GTT  | CTT  | CTT   | CCT   | CTT   | CTTCCI  | 'ŒG  | CGT  | GTG  | GCT   | GG"  | PAGC  | CTA'   | rgg  | CGT  | 'GG  |
| K   | D V  | F  | F  | F   | Ъ   | F   | F L   | G  | V  | W  | L   | V  | A   | Y  | G  | V  | A  |
|   | 2890   |  |  |   |   |   | 2910  |  |  |  |   |  | 2930  | 0  |  |  |  |
| CCAC  | GGAGGG   | GCT  | CCT  | GAG   | GCC   | ACG   | GGACAG  | TGA  | CTT  | ccc  | AAG   | TA?  | CCT   | 3CG(   | CCG  | CGT  | CT   |
| T   | E G  | L  | L  | R   | P   | R   | D S   | Ø  | F  | P  | S   | I  | L   | R  | R  | ٧  | F  |
|   | 2950   |  |  |   |   |   | 2970  |  |  |  |   |  | 2990  | )  |  |  |  |
| TCTA  | CCGTCC   | CTA  | CCT  | GCA   | GAT   | CTT   | CGGGCZ  | GAT  | TCC  | CCA  | GGA   | GGZ  | ACAT  | 3GA  | CGT  | GGC  | CC   |
| Y   |  |  | L  | Q   | I   | F   | G Q   | I  | P  | Q  | E   | D  | M   | D  | v  | A  | L  |
|   | R P  | Y  |  |   |   |   |   |  |  |  |   |  |   |  | •  | A  |  |
|   | R P<br>3010  |  | -  |   |   |   | 3030  |  |  |  |   |  | 3050  | 0  | •  | A  | _  |
| TCAT  | 3010   |  |  | CTG   | CTC   | GTC   |   | :CGG                                       | CTT  | CTG  | GGC   | AC   |   |  |  |  | ccc  |
|   | 3010<br>GGAGCA   | .CAG   | CAA  |   | -   |   | GGAGCC  |  |  |  |   |  | ACCC.   | rcc:   | TGG  | GGC  |  |
| TCAI<br>M   | 3010<br>GGAGCA<br>E H  | CAG<br>S   |  | CTG<br>C                                    | -   |   | GGAGCO<br>E P   |  |  | ctg<br>W   | GGC<br>A                                    |  | ACCC.   | rcc:<br>P  |  | GGC  | ccc<br>Q   |
| M   | 3010<br>GGAGCA<br>E H<br>3070  | CAG<br>S   | CAA<br>N   | С   | S   | s   | GGAGCO<br>E P<br>3090   | G  | F  | W  | A   | H  | ACCC:<br>P<br>311   | P<br>D   | TGG<br>G                                       | GGC<br>A   | Q  |
| M<br>AGGC   | 3010<br>GGAGCA<br>E H<br>3070<br>GGGGCAC   | CAG<br>S<br>CTG  | CAA<br>N<br>CGT                                    | c<br>CTC                                    | S<br>CCA  | s<br>gta  | GGAGCO<br>E P<br>3090<br>TGCCAF   | G<br>ACTG                                  | F<br>GCT   | W<br>GGT   | a<br>GGT                                    | H<br>GC  | ACCC<br>P<br>3110<br>IGCT   | P<br>O<br>CCT  | TGG<br>G<br>CGT                                | GGC<br>A<br>CAT  | Q  |
| M   | 3010<br>EGAGCA<br>E H<br>3070<br>EGGGCAC<br>G T  | CAG<br>S<br>CTG  | CAA<br>N   | С   | S   | s   | GGAGCO<br>E P<br>3090<br>TGCCAF<br>A N  | G  | F  | W  | a<br>GGT                                    | H<br>GC  | ACCC:<br>P<br>3110<br>IGCTO<br>L  | P<br>O<br>CCT  | TGG<br>G                                       | GGC<br>A<br>CAT  | Q  |
| M<br>AGGC<br>A                                      | 3010<br>GGAGCA<br>E'H<br>3070<br>GGGCAC<br>G T<br>3130   | CAG<br>S<br>CTG<br>C   | CAA<br>N<br>CGT<br>V                               | C<br>CTC<br>S                               | S<br>CCA<br>Q                                   | S<br>GTA<br>Y   | GGAGCO<br>E P<br>3090<br>TGCCAF<br>A N<br>3150  | G<br>ACTG<br>W                             | F<br>GCT<br>L  | W<br>GGT<br>V  | A<br>GGT<br>V                               | H<br>GC:<br>L  | ACCC<br>P<br>3110<br>FGCTO<br>L<br>3170   | P<br>O<br>CCT<br>L   | TGG<br>G<br>CGT<br>V                           | GGC<br>A<br>CAT  | Q<br>CT<br>F   |
| AGGC<br>A   | 3010<br>GGAGCA<br>E H<br>3070<br>GGGCAC<br>G T<br>3130   | CAG<br>S<br>CTG<br>C   | CAA<br>N<br>CGT<br>V<br>CAA                        | C<br>CTC<br>S<br>CAT                        | S<br>CCA<br>Q<br>CCT                            | S<br>GTA<br>Y<br>GCT  | GGAGCO E P 3090 TGCCAF A N 3150 GGTCAF  | G<br>LCTG<br>W<br>LCTT                     | F<br>GCT<br>L<br>GCT   | W<br>GGT<br>V<br>CAT   | A<br>GGT<br>V<br>TGC                        | H<br>GC:<br>L<br>CA:   | ACCC: P 311( FGCT) L 317( FGTT)   | P<br>O<br>CCT<br>L<br>O<br>CAG   | TGG<br>G<br>CGT<br>V                           | GGC<br>A<br>CAT  | Q<br>TCT<br>F  |
| M<br>AGGC<br>A                                      | 3010<br>GGAGCA<br>E'H<br>3070<br>GGGCAC<br>G T<br>3130   | CAG<br>S<br>CTG<br>C   | CAA<br>N<br>CGT<br>V                               | C<br>CTC<br>S                               | S<br>CCA<br>Q                                   | S<br>GTA<br>Y<br>GCT  | GGAGCO E P 3090 TGCCAF A N 3150 GGTCAF V N  | G<br>LCTG<br>W<br>LCTT                     | F<br>GCT<br>L  | W<br>GGT<br>V  | A<br>GGT<br>V                               | H<br>GC:<br>L  | ACCC: P 3110 FGCT0 L 3170 FGTT0 F   | P<br>O<br>CCT<br>L<br>O<br>CAG   | TGG<br>G<br>CGT<br>V                           | GGC<br>A<br>CAT  | Q<br>CT<br>F   |
| AGGC<br>A   | 3010<br>GGAGCA<br>E H<br>3070<br>GGGCAC<br>G T<br>3130   | CAG<br>S<br>CTG<br>C<br>C<br>GGC   | CAA<br>N<br>CGT<br>V<br>CAA                        | C<br>CTC<br>S<br>CAT                        | S<br>CCA<br>Q<br>CCT                            | S<br>GTA<br>Y<br>GCT  | GGAGCO E P 3090 TGCCAF A N 3150 GGTCAF  | G<br>LCTG<br>W<br>LCTT                     | F<br>GCT<br>L<br>GCT   | W<br>GGT<br>V<br>CAT   | A<br>GGT<br>V<br>TGC                        | H<br>GC:<br>L<br>CA:   | ACCC: P 311( FGCT) L 317( FGTT)   | P<br>O<br>CCT<br>L<br>O<br>CAG   | TGG<br>G<br>CGT<br>V                           | GGC<br>A<br>CAT  | Q<br>TCT<br>F  |
| AGGC<br>A<br>TCCI<br>L                              | 3010<br>CGGAGCA<br>E' H<br>3070<br>CGGGCAC<br>G T<br>3130<br>CGCTCGT<br>L V  | CAG<br>S<br>CTG<br>C<br>C<br>GGC<br>A  | CAA<br>N<br>CGT<br>V<br>CAA<br>N                   | CTC<br>S<br>CAT                             | S<br>CCA<br>Q<br>CCT<br>L                       | S<br>GTA<br>Y<br>GCT<br>L                                     | GGAGCO E P 3090 TGCCAF A N 3150 GGTCAF V N 3210   | G<br>W<br>ACTT<br>L                        | F<br>GCT<br>L<br>GCT<br>L  | W<br>GGT<br>V<br>CAT   | A<br>GGT<br>V<br>TGC<br>A                   | H<br>GC:<br>L<br>CA!   | ACCCI<br>P<br>311(<br>FGCT)<br>L<br>317(<br>FGTT)<br>F  | POCTO  | TGG<br>G<br>CGT<br>V<br>TTA                    | GGC<br>A<br>CAT<br>I<br>CAC  | Q<br>F<br>F<br>CAT<br>F  |
| M AGGC A TCCT L                                     | 3010<br>EGGAGCA<br>E' H<br>3070<br>EGGGCAC<br>G T<br>3130<br>EGCTCGT<br>L V<br>3190                                | CAG<br>CTG<br>C<br>C<br>CGGC<br>A  | CAA<br>N<br>CGT<br>V<br>CAA<br>N                   | CTC<br>S<br>CAT<br>I                        | S<br>CCA<br>Q<br>CCT<br>L                       | S<br>GTA<br>Y<br>GCT<br>L<br>CGA                              | GGAGCO E P 3090 TGCCAF A N 3150 GGTCAF V N 3210   | G<br>W<br>ACTT<br>L                        | GCT<br>L<br>GCT<br>L<br>GAA  | W<br>CAT<br>I  | A<br>GGT<br>V<br>TGC<br>A                   | GC:<br>L<br>CA:<br>M   | ACCC<br>P<br>3110<br>FGCT0<br>L<br>3170<br>FGTT0<br>F<br>3230   | P<br>CCT<br>L<br>CAG<br>S<br>CCG   | TGG G CGT V TTA Y CCT                          | CAC  | Q<br>F<br>F<br>CAT<br>F  |
| M AGGC A TCCT L                                     | 3010 GGAGCA E' H 3070 GGGGCAC G T 3130 GGCTCGT L V 3190 GCAAAGT  | CAG<br>S<br>CTG<br>C<br>C<br>GGC<br>A  | CAA<br>N<br>CGT<br>V<br>CAA<br>N                   | CTC<br>S<br>CAT<br>I                        | S<br>CCA<br>Q<br>CCT<br>L                       | S<br>GTA<br>Y<br>GCT<br>L<br>CGA                              | GGAGCO E P 3090 TGCCAF A N 3150 GGTCAF V N 3210   | G<br>W<br>ACTT<br>L                        | GCT<br>L<br>GCT<br>L<br>GAA  | W<br>CAT<br>I  | A<br>GGT<br>V<br>TGC<br>A                   | GC:<br>L<br>CA:<br>M   | ACCC<br>P<br>3110<br>FGCT0<br>L<br>3170<br>FGTT0<br>F<br>3230   | P<br>CCT<br>L<br>CAG<br>S<br>CCG   | TGG G CGT V TTA Y CCT                          | CAC  | Q<br>F<br>F<br>CAT<br>F  |
| M AGGC A TCCT L TCGG                                | 3010 GGGAGCA E' H 3070 GGGGCAC G T 3130 GGCTCGT L V 3190 GCAAAGT K V 3250  | CAG<br>S<br>CTG<br>C<br>C<br>GGC<br>A  | CAA<br>N<br>CGT<br>V<br>CAA<br>N<br>GGG            | CTC<br>S<br>CAT<br>I<br>CAA<br>N            | S<br>CCA<br>Q<br>CCT<br>L<br>CAG                | S<br>GTA<br>Y<br>GCT<br>L<br>CGA<br>D                         | GGAGCO E P 3090 TGCCAF A N 3150 GGTCAF V N 3210 TCTCTF L Y 3270   | G<br>W<br>ACTT<br>L<br>CTG<br>W            | F<br>GCT<br>L<br>GCT<br>L<br>GAA                                       | W<br>V<br>CAT<br>I<br>GGC  | A<br>V<br>TGC<br>A<br>CGCA                  | H<br>GCCAL<br>M<br>MGCCC<br>R  | ACCC: P 3110 FGCT0 L 3170 FGTT0 F 3230 GTTA0 Y 3290   | POCCTOR SOCOGO R   | TGG G CGT V TTA Y CCT                          | CAC<br>T   | Q<br>F<br>EAT<br>F<br>FCC<br>R   |
| AGGCA A TCCT L TCGG                                 | 3010 GGAGCA E' H 3070 GGGCAC G T 3130 GGCTCGT L V 3190 GCAAAGT K V 3250 ATTCCA                                     | CAG S CTG C A CAG C C C C C C C C C C C C C C C  | CAA<br>N<br>CGT<br>V<br>CAA<br>N<br>GGG<br>G       | CTC<br>S<br>CAT<br>I<br>CAA<br>N            | S CCCA Q CCT L CAG S CGC                        | S<br>GTA<br>Y<br>GCT<br>L<br>CGA<br>D                         | GGAGCC E P 3090 TGCCAF A N 3150 GGTCAF V N 3210 TCTCTA L Y 3270 GGCCCC  | G<br>W<br>ACTT<br>L<br>ACTG<br>W           | F<br>GCT<br>L<br>GGAA<br>K   | W<br>CAT<br>I<br>GGC<br>A  | A GGT V TGC A GCA Q CGT                     | H GCA: M GCCA: R CCA:  | ACCC: P 3110 FGCT0 L 3170 FGTTC F 3230 FTTAC Y 3290 FCTC  | P 0 CCT L 0 CAG S 0 CCG R 0 CCG  | TGG G CGT V TTA Y CCT L                        | GGGC<br>A<br>CAT<br>CAC<br>T   | Q<br>F<br>F<br>CAT<br>F<br>FCC<br>R  |
| AGGCA A TCCT L TCGG                                 | 3010 GGAGCA E' H 3070 GGGCAC G T 3130 GGCTCGT L V 3190 GCAAAGT K V 3250 ATTCCA                                     | CAG CTG CGGC A ACACA Q ACTC  | CAA<br>N<br>CGT<br>V<br>CAA<br>N<br>GGG<br>G       | CTC<br>S<br>CAT<br>I<br>CAA<br>N            | S CCCA Q CCT L CAG S CGC                        | S<br>GTA<br>Y<br>GCT<br>L<br>CGA<br>D                         | GGAGCC E P 3090 TGCCAF A N 3150 GGTCAF V N 3210 TCTCTF L Y 3270 GGCCCC A P  | G<br>W<br>ACTT<br>L<br>ACTG<br>W           | F<br>GCT<br>L<br>GGAA<br>K   | W<br>CAT<br>I<br>GGC<br>A  | A GGT V TGC A GCA Q CGT                     | H GCA: M GCCA: R CCA:  | ACCC: P 3110 FGCT0 L 3170 FGTTC F 3230 FGTTA Y 3290 FCTC0   | POCCTOR COGGER COCCA H   | TGG G CGT V TTA Y CCT L                        | GGGC<br>A<br>CAT<br>CAC<br>T   | Q<br>F<br>F<br>CAT<br>F<br>FCC<br>R  |
| M AGGC A TCCT L TCGG G GGGA                         | 3010 GGAGCA E'H 3070 GGGCAC G T 3130 GGCTCGT L V 3190 GCAAAGT K V 3250 ATTCCA                                      | CAG CTG CAG CAG CAG CAG CAG CAG CAG CAG CAG CA   | CAA<br>N<br>CGT<br>V<br>CAA<br>N<br>GGG<br>G       | C CTC S CAT I CAA N GCC P                   | S CCA CCT L CAG S CCGC A                        | S<br>GTA<br>Y<br>GCT<br>L<br>CGA<br>D<br>GCT<br>L             | GGAGCC E P 3090 TGCCAP A N 3150 GGTCAP V N 3210 TCTCTP L Y 3270 GGCCCC A P 3330   | G<br>ACTG<br>W<br>ACTT<br>L<br>ACTG<br>W   | GCT<br>L<br>GCT<br>L<br>GAA<br>K                                       | W<br>CAT<br>I<br>GGC<br>A  | A GGT V TGC A GCA Q CGT V                   | H<br>GCCAN<br>M<br>GCCAN<br>R  | ACCC:   | POCCE POCCE CAG CCAG CCCG R CCCG H   | TGG G CGT V TTA Y CCT L                        | GGGC<br>A<br>CAT<br>CAT<br>CAT<br>I  | Q<br>FCT<br>F<br>CAT<br>F<br>CCC<br>R  |
| M AGGC A TCCT L TCGG G GGGA E TCCT                  | 3010 CGGAGCA E' H 3070 CGGCACA G T 3130 CGCCAAAGT K V 3250 ATTCCA F H 3310 CGCCCAG                                 | CAG CTG CAG CCTG A CACA CACA CACA CACA C   | CAA N CGT V CAA N GGG G TCG                        | C CTC S CAT I CAA N GCC P GTG               | S CCA Q CCT L CAG S CGC A                       | S GTA Y GCT L CGA D GCT L                                     | GGAGCC E P 3090 TGCCAP A N 3150 GGTCAP V N 3210 TCTCTAP L Y 3270 GGCCCC A P 3330 ACCCCC   | G<br>ACTG<br>W<br>ACTT<br>L<br>ACTG<br>W   | F<br>GCT<br>L<br>GGAA<br>K<br>CCTT<br>F                                | W CGGT V CAT I GGCC A TAT I  | A GGT V TGC A GCA Q CGT V                   | H GCC L CAS M GCC R CAS CCAS CCAS CCAS CCAS CCAS CCAS C  | ACCC: P 3111 FGCT0 L 3170 FGTT0 F 3230 FTTAV Y 3290 FCTC0 S 3355  | POCCES CCAG R CCCG R CCCG R CCCG R CCCCAC  | TGG G CGT Y TTA Y CCT L CTT L                  | GGGC<br>A<br>CAT<br>I<br>CAC<br>T<br>CAT<br>I  | Q<br>TCT<br>F<br>CAT<br>F<br>CCC<br>R<br>GCC<br>L  |
| M AGGC A TCCT L TCGG G GGGA                         | 3010 GGAGCA E H 3070 GGGCAC G T 3130 GGCTCGT K V 3250 ATTCCA F H 3310 GGCTCAG                                      | CAGC CCTG CCTG A CACA CACA CACA CACA CAC   | CAA N CGT V CAA N GGG G TCG                        | C CTC S CAT I CAA N GCC P GTG               | S CCA Q CCT L CAG S CGC A                       | S GTA Y GCT L CGA D GCT L                                     | GGAGCC E P 3090 TGCCAP A N 3150 GGTCAP V N 3210 TCTCTP L Y 3270 GGCCCC A P 3330 ACCCCCC   | G<br>ACTG<br>W<br>ACTT<br>L<br>ACTG<br>W   | F<br>GCT<br>L<br>GGAA<br>K<br>CCTT<br>F                                | W CGGT V CAT I GGCC A TAT I  | A GGT V TGC A GCA Q CGT V                   | H GCC L CAS M GCC R CAS CCAS CCAS CCAS CCAS CCAS CCAS C  | P 3110 FGCTC L 3177 FGTTC F 3230 SGTTA Y 3299 SGTTCTC S 3355 CCCCCCC S  | POCCE<br>POCCE<br>CAG<br>SOCCE<br>ROCCE<br>ROCCE<br>POCCE<br>P   | TGG G CGT Y TTA Y CCT L CTT L                  | GGGC<br>A<br>CAT<br>CAC<br>T<br>CAT<br>I   | Q<br>TCT<br>F<br>CAT<br>F<br>CCC<br>R<br>GCC<br>L  |
| M AGGGA TCCT L TCGG G GGGA E TCCT L                 | 3010 GGAGCA E' H 3070 GGGCAC G T 3130 GCTCGT L V 3190 GCAAAGT K V 3250 AATTCCA F H 3310 GGCTCAG                    | CAG CTG C A CACA CACA CACA CACA CACA CAC   | CAA N CGT V CAA N GGG G TCG R ATT                  | C CTC S CAT I CAA N GCC P GTG C             | S CCA CAG                                       | S<br>GTA<br>Y<br>GCT<br>L<br>CGA<br>D<br>GCT<br>L<br>GCG<br>R | GGAGCC E P 3090 TGCCAP A N 3150 GGTCAP V N 3210 TCTCTP L Y 3270 GGCCCC A P 3330 ACCCCC P R 3390   | G W ACTT L ACTG W CGCC P GGAG              | GCT<br>L<br>GCT<br>L<br>GAA<br>K<br>CCTT<br>F                          | W CGGT V CAT I GGC A TAT I CCA   | A GGT V TGC A GCA Q CGT V GGCO P            | H<br>GCC<br>L<br>CCA:<br>M<br>GCC<br>R<br>CCA:<br>I  | 3110<br>FGCT0<br>L<br>3170<br>FGTT0<br>F<br>3230<br>FGTTA<br>Y<br>3290<br>FCCCCCCC<br>S<br>3355<br>CCCCCCC<br>S | POCTOR DOCOR | TGG G CGT V TTA Y CCT L CTT L                  | GGGGGA  CAT  CAT  CAT  CAT  CAT  CAT  CA   | Q<br>FCT<br>F<br>CAT<br>F<br>FCC<br>R<br>GCC<br>L  |
| M AGGGA A TCCT L TCGG G G GGGA E TCCT L             | 3010 CGGAGCA E' H 3070 CGGCACA G T 3130 CGCTCGT L V 3190 CGCAAAGT K V 3250 ATTCCA L R 3370 L R                     | CAG S CTG C A ACA Q CTC S GGCA GGCA  | CAA  CGT  V  CAA  N  GGG  G  TCG  R  ATT  TTA      | C CTC S CAT I CAA N GCC P GTG C CCT         | S CCA Q CCT L CAG S CGC A CAG R TTC             | S GTA Y GCT L CGA D GCT L GCG R TAA                           | GGAGCC E P 3090 TGCCAF A N 3150 GGTCAF V N 3210 TCTCTF L Y 3270 GGCCCC A P 3330 ACCCCCC P R 3390 GGAAGC                                 | G ACTG W ACTT L ACTG W CGCC P GGAG S CCGA  | GCT L GGAA K CTT F CCCC P GCG  | W CGGT V CAT I GGC A TAT I CCA Q GGAA  | A GGT V GGCA P GCT                          | H GCA: M GCCA: T GCCA: | ACCC: P 3110 L 3170 F 3170 F 3230 F 3290 F TCTO S 3355 CCCTCC S 3410  | POCCE POCCE CAG CCG R CCCA CCCA H CCCC P CCCC CCG CCCC CCCC CCCC CCCC C  | TGG G CGT V TTA Y CCT L CTT L GGC A            | GGGGGA  CAT  CAT  CAT  CAT  CAT  CAT  ATC  | Q<br>CT<br>F<br>CAT<br>F<br>CCC<br>R<br>SCCC<br>L  |
| M AGGGA A TCCT L TCGG G G GGGA E TCCT L             | 3010 GGAGCA E' H 3070 GGGCAC G T 3130 GCTCGT L V 3190 GCAAAGT K V 3250 AATTCCA F H 3310 GGCTCAG                    | CAG S CTG C A ACA Q CTC S GGCA GGCA  | CAA  CGT  V  CAA  N  GGG  G  TCG  R  ATT  TTA      | C CTC S CAT I CAA N GCC P GTG C CCT         | S CCA Q CCT L CAG S CGC A CAG R TTC             | S GTA Y GCT L CGA D GCT L GCG R TAA                           | GGAGCC E P 3090 TGCCAF A N 3150 GGTCAF V N 3210 TCTCTF L Y 3270 GGCCCC A P 3330 ACCCCCC P R 3390 GGAAGC                                 | G ACTG W ACTT L ACTG W CGCC P GGAG S CCGA  | GCT L GGAA K CTT F CCCC P GCG  | W CGGT V CAT I GGC A TAT I CCA Q GGAA  | A GGT V GGCA P GCT                          | H GCA: M GCCA: T GCCA: | ACCC: P 3110 L 3170 F 3170 F 3230 F 3290 F TCTO S 3355 CCCTCC S 3410  | POCCE POCCE CAG CCG R CCCA CCCA H CCCC P CCCC CCG CCCC CCCC CCCC CCCC C  | TGG G CGT V TTA Y CCT L CTT L GGC A            | GGGGGA  CAT  CAT  CAT  CAT  CAT  CAT  ATC  | Q<br>CT<br>F<br>CAT<br>F<br>CCC<br>R<br>SCCC<br>L  |
| M AGGGA TCCT L TCGG G GGGA E TCCT L AGCA            | 3010 CGGAGCA E' H 3070 CGGCACA G T 3130 CGCTCGT K V 3250 ATTCCA F H 3310 CGCTCAG L R 3370 TTTCCG F R               | CAG S CTG C C C C C C C C C C C C C C C C C C  | CAA N CGT V CAA N GGG G TCG R ATT L TTA            | C CTC S CAT I CAA N GCC P GTG C CCT L       | S CCA Q CCT L CAG S CGC A CAG R TTC             | S GTA Y GCT L CGA D GCT L GCG R TAA                           | GGAGCC E P 3090 TGCCAF A N 3150 GGTCAF V N 3210 TCTCTF L Y 3270 GGCCCC A P 3330 ACCCCCC P R 3390 GGAAGC E A 3450                        | G ACTG W ACTT L ACTG W CGCC P CGGAG S CCGA | F<br>GCT<br>L<br>GAAA<br>K<br>CTT<br>F<br>CCC<br>P<br>GCG<br>R         | W<br>CAT<br>1<br>GGC<br>A<br>TAT<br>1<br>CCA<br>Q  | A GGT V TGC A GCA Q CGT V GCC P             | H GCC L CAN M GCC R CAN CCAN CCAN CCAN CCAN CCAN CCAN C  | ACCC: P 3111 FGCTT L 317 FGTTT F 323 Y 329 FCTCC S 335 CCCTCC S 3410 TAACC T                                    | POCCE<br>POCCE<br>CAG<br>SCCG<br>ROCCA<br>HOCCC<br>POCCE<br>POCCE<br>WOO   | TGG G CGT Y TTA Y CCT L CTT L GGC A GGA E      | CAT  | Q<br>TCT<br>F<br>CCC<br>R<br>CCC<br>L<br>CCG<br>E  |
| M AGGGA TCCT L TCGG G GGGA E TCCT L AGCA            | 3010 GGAGCA E' H 3070 GGGCAC G T 3130 GCTCGT L V 3190 GCAAAGT K V 3250 AATTCCA L R 3370 ATTTCCG F R                | CAG S CTG C C C C C C C C C C C C C C C C C C  | CAA N CGT V CAA N GGG G TCG R ATT L TTA            | C CTC S CAT I CAA N GCC P GTG C CCT L       | S CCA Q CCT L CAG S CGC A CAG R TTC             | S GTA Y GCT L CGA D GCT L GCG R TAA                           | GGAGCC E P 3090 TGCCAF A N 3150 GGTCAF V N 3210 TCTCTF L Y 3270 GGCCCC A P 3330 ACCCCCC P R 3390 GGAAGC E A 3450                        | G ACTG W ACTT L ACTG W CGCC P CGGAG S CCGA | F<br>GCT<br>L<br>GAAA<br>K<br>CTT<br>F<br>CCC<br>P<br>GCG<br>R         | W<br>CAT<br>1<br>GGC<br>A<br>TAT<br>1<br>CCA<br>Q  | A GGT V TGC A GCA Q CGT V GCC P             | H GCC L CAN M GCC R CAN CCAN CCAN CCAN CCAN CCAN CCAN C  | ACCC: P 3111 FGCTT L 317 FGTTT F 323 Y 329 FCTCC S 335 CCCTCC S 3410 TAACC T                                    | POCCE<br>POCCE<br>CAG<br>SCCG<br>ROCCA<br>HOCCC<br>POCCE<br>POCCE<br>WOO   | TGG G CGT Y TTA Y CCT L CTT L GGC A GGA E      | CAT  | Q<br>TCT<br>F<br>CCC<br>R<br>CCC<br>L<br>CCG<br>E  |
| M AGGGA A TCCT L TCGG G G GGGA E TCCT L AGCA H TGCA | 3010 CGGAGCA E' H 3070 CGGCACA G T 3130 CGCTCGT K V 3250 ATTCCA G T 3370 CGCTCAG F H 3370 CGCTCAG F R 3430 TTAAGGA | CAG S CCTG C C C C C C C C C C C C C C C C C   | CAA N CGT V CAA N GGG G TCG R ATT L TTA Y CTT      | C CTC S CAT I CAA N GCC P GTG C CCT L       | S CCCA Q CCT L CAG S CGC A CAG R TTC S GCT      | S GTA Y GCT L CGA D GCT L GCG R TAA K                         | GGAGCC E P 3090 TGCCAF A N 3150 GGTCAF V N 3210 TCTCTF L Y 3270 GGCCCC A P 3330 ACCCCCC P R 3390 GGAAGC E A 3450 ACCCCCC                | G ACTO W ACTT L ACTG W CGCC P CGGAG S CCGA | GCT<br>L<br>GCT<br>L<br>GAA<br>K<br>CCTT<br>F<br>CCCC<br>P<br>GCG<br>R | W CGGT V CAT I GGCC A CCCA CCCA CCCA K CCCA K CCCA K CCCA CCCCA CCCA CCCA CCCA CCCCA CCCA CCCA CCCA CCCCA CCCCCA CCCCA CCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCC | A GGT V TGC A GCA Q GGT V GCC P GCC L GCG   | H GCA: M GCA: R GCA: SGT( SGT( SGT( SGT( SGGG) SGGG)   | ACCC: P 3111 FGCTT L 317 FGTTT F 323 Y 329 FCTTO S 335 CCCTCC S 3411 TAACC T 347 AGAGG                          | POCT POCCA ROCCA HOCCA POCCA WOCCA W | TGG G CGT V TTA Y CCT L CTT L GGC A GGA E      | CAT  CAT  CAT  CAT  CAT  CAT  CAT  CAT   | Q CT F CAT F CC R R SCC L CC E CC E CC E CC E CC E CC E CC   |
| M AGGGA A TCCT L TCGG G G GGGA E TCCT L AGCA H TGCA | 3010 CGGAGCA E' H 3070 CGGCACA G T 3130 CGCTCGT K V 3250 ATTCCA F H 3310 CGCTCAG L R 3370 TTTCCG F R               | CAG CTG CC   | CAA N CGT V CAA N GGG G TCG R ATT L TTA Y CTT      | C CTC S CAT I CAA N GCC P GTG C CCT L       | S CCCA Q CCT L CAG S CGC A CAG R TTC S GCT      | S GTA Y GCT L CGA D GCT L GCG R TAA K                         | GGAGCC E P 3090 TGCCAP A N 3150 GGTCAP V N 3210 TCTCTA L Y 3270 GGCCCC A P 3330 ACCCCC P R 3390 GGAAGC E A 3450 ACGCGC R A              | G ACTO W ACTT L ACTG W CGCC P CGGAG S CCGA | GCT<br>L<br>GCT<br>L<br>GAA<br>K<br>CCTT<br>F<br>CCCC<br>P<br>GCG<br>R | W CGGT V CAT I GGCC A CCCA CCCA CCCA K CCCA K CCCA K CCCA CCCCA CCCA CCCA CCCA CCCCA CCCA CCCA CCCA CCCCA CCCCCA CCCCA CCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCC | A GGT V TGC A GCA Q GGT V GCC P GCC L GCG   | H GCA: M GCA: R GCA: SGT( SGT( SGT( SGT( SGGG) SGGG)   | ACCC: P 3111 FGCTT L 317 FGTTT F 323 Y 329 FCTTO S 335 CCCTCC S 3411 TAACC T 347 AGAGG                          | POCTOR DOCAGO DO | TGG G CGT V TTA Y CCT L CTT L GGC A GGA E      | CAT  CAT  CAT  CAT  CAT  CAT  CAT  CAT   | Q CT F CAT F CC R R SCC L CC E CC E CC E CC E CC E CC E CC   |
| M AGGGA A TCCT L TCGG G GGGA E TCCT L AGCA H        | 3010 GGAGCA E' H 3070 GGGCACA G T 3130 GCTCGT K V 3250 ATTCCA GT 1 R 3370 TTTCCG F R 3430 TTAAGGA K E 3490         | CAG<br>S<br>CCTG<br>CC<br>CAG<br>CACA<br>CCTC<br>S<br>GGCA<br>V  | CAA N CGT V CAA N GGG G TCG R TTA TTA Y CTT F      | C CTC S CAT I CAA N GCC P GTG C C CT L      | S CCA Q CAG S CAG R TTC S GCT L                 | S GTA Y GCT L CGA D GCT L GCG R TAA K GGC A                   | GGAGCC E P 3090 TGCCAF A N 3150 GGTCAF V N 3210 TCTCTF L Y 3270 GGCCCC A P 3330 ACCCCCC P R 3390 GGAAGC E A 3450 ACGCGC R A 3510        | G CTG W CCTG P CCCGA E R                   | F GCT L GCAA K CCTT F GCG P GCG R GGAA                                 | W GGT V CAT I GGC A CCAA CCAA K CAA K  | A GGT V TGC A GCA CGT V GCC P GCT L GCG R   | H GCC  | ACCC: P 3111 FGCTT L 317 FGTTT F 323 Y 329 FCTCC S 3350 CCCCC T 347 AGAGG S 353                                 | POCT POCCA P | TGG G CGT Y TTA Y CCT L GGC A GGA E CTC        | GGGC<br>A<br>CAT<br>I<br>CAT<br>I<br>CCT<br>I<br>CCT<br>I<br>CCT<br>I  | Q CT F CCC R SCC L CCC E |
| M AGGGA A TCCT L TCGG G G GGGA H TCCT H GTCT        | 3010 GGAGCA E' H 3070 GGGCAC G T 3130 GCTCGT L V 3190 GCAAAGT K V 3250 ATTTCCA L R 3370 ATTTCCC F R 3430 K E       | CAG CTG CGCA CACACA CACACACA CACACA CACACA CACACACA CACACA CACACA CACACA CACACA | CAA  CGT  V  CAA  N  GGG  TCG  R  TTA  TTA  Y  GTC | C CTC S CAT I CAA N GCC P GTG C CCT L TCT L | S CCA Q CCT L CAG S CAG R CAG R TTC S GCT L GAA | S GTA Y GCT L CGA D GCT L GCG R TAA K GGC A                   | GGAGCC E P 3090 TGCCAF A N 3150 GGTCAF V N 3210 TCTCTF L Y 3270 GGCCCC A P 3330 ACCCCCC P R 3390 GGAAGC E A 3450 ACGCGC R A 3510 GGACTT | G CTG W CCTG W CCCG P CCCGA E R CCCGA R    | F GCT L GAA K CTT F GCC P GCG R GGA ACT                                | W GGGT V CAT I GGGC A CCCA CCCA K CCAA K CCAA  | A GGT V TGC A GCA Q CGT V GCC P GCT L GCG R | H GCC L GCC R GCC L GCC E GCC E  | ACCC: P 3111 FGCTT L 317 FGTTT F 323 Y 329 FCTTO S 335 CCCTCC S 3411 TAACC T 347 AGAGG S 353                    | POCT POCCA P | TGGG G CGT Y TTA Y CCT L CTT L GGC A CTC S CAT | GGGC A CAT I | Q CT F CCC R SCC L CCG E CCG R SCC C R |

Fig. 9 / continua 13

3570 3590 AGTACGAACAGCGCCTGAAAGTGCTGGAGCGGGAGGTCCAGCAGTGTAGCCGCGTCCTGG Y E Q R L K V L B R E V Q Q C S R V L G 3630 3650 GGTGGGTGGCCGAGGCCCTGAGCCGCTCTGCCTTGCTGCCCCCAGGTGGGCCGCCACCCC WVAEALSRSALLPPGGPPPP 3670 3690 3710  $\tt CTGACCTGCCTGGGTCCAAAGACTGAGCCCTGCTGGCGGACTTCAAGGAGAAGCCCCCAC$ D L P G S K D \* 3730 3750 3770 AGGGGATTTTGCTCCTAGAGTAAGGCTCATCTGGGCCTCGGCCCCCGCACCTGGTGGCCT 3810 3830 TGTCCTTGAGGTGAGCCCCATGTCCATCTGGGCCACTGTCAGGACCACCTTTGGGAGTGT 3850 3870 3890 CATOCTTACAAACCACAGCATGCCCGGCTCCTCCCAGAACCAGTCCCAGCCTGGGAGGAT 3910 3930 3950 CAAGGCCTGGATCCCGGGCCGTTATCCATCTGGAGGCTGCAGGGTCCTTGGGGTAACAGG 3970 3990 4010 GACCACAGACCCCTCACCACTCACAGATTCCTCACACTGGGGAAATAAAGCCATTTCAGA 4030 **GGAAAAAAAAAAAAAA** 

MVVPEKEQSWIPKIFKKKTCTTFIVDSTDPGGTLCQCGRPRTAHPAVAMEDAFGAAVVTVWDSDAHTTEKPTDAYELDFTGAGRKE SNFLRLSDRTDPAAVYSLVTRTWGFRAPNLVVSVLGGSGGFVLQTWLQDLLRRGLVRAAQSTGAWIVTGGLHTGIGRHVGVAVRDH QMASTGGTKVVAMGVAPWGVVRNRDTLINPKGSFPARYRWRGDPEDGVQFPLDYNYSAFFLVDDGTHGCLGGENRFRLRLESYISQ QKTGVGGTGIDIPVLLLLIDGDEKMLTRIENATQAHVPCLLVAGSRGLGNPGGTLEAHLAQDGDHKANQSTNQLLLPKDLSLQPVE SIDRKTLQSYSERLAVAWNRVDIAQSELFRGDIQWRSFHLEASLMDALLNDRPEFVRLLISHGLSLGHFLTPMRLAQLYSAAPSNS LIRNLLDQASHSAGTKAPALKGGAAELRPPDVGHVLRMLLGKMCAPRYPSGGAWDPHPGQGFGESMYLLSDKATSPLSLDAGLGQA PWSDLLLWALLLNRAQMAMYFWEMGSNAVSSALGACLLLRVMARLEPDAEEAARRKDLAFKFEGMGVDLFGECYRSSEVRAARLLL RRCPLWGDATCLQLAMQADARAFFAQDGVQSLLTQKWWGDMASTTPIWALVLAFFCPPLIYTRLITFRKSEEEPTREELEFDMDSV INGEGPVGTADPAEKTPLGVPRQSGRPGCCGGRCGGRRCLRRWFHFWGVPVTIFMGNVVSYLLFLLLFSRVLLVDFQPAPPGSLEL LLYFWAFTLLCEELRQGLSGGGGSLASGGPGPGHASLSQRLRLYLADSWNQCDLVALTCFLLGVGCRLTPGLYHLGRTVLCIDFMV FTVRLLHIFTVNKQLGPKIVIVSKMMKDVFFFLFFLGVWLVAYGVATEGLLRPRDSDFPSILRRVFYRPYLQIFGQIPQEDMDVAL MEHSNCSSEPGFWAHPFGAQAGTCVSQYANWLVVLLLVIFLLVANILLVNLLIAMFSYTFGKVQGNSDLYWKAQRYRLIREFHSRP ALAPPFIVISHLRLLLRQLCRRPRSPQPSSPALEHFRVYLSKEAERKLLTWESVHKENFLLARARDKRESDSERLKRTSQKVDLAL KQLGHIREYEQRLKVLEREVQQCSRVLGWVAEALSRSALLPPGGPPPPDLPGSKD

в.)

|      |      | 10   |     |     |     |      |         | •   | 30  | 50  |     |     |      |     |     |      |      |      |     |
|------|------|------|-----|-----|-----|------|---------|-----|-----|-----|-----|-----|------|-----|-----|------|------|------|-----|
| ATC  | 'AA' | rggc | GGT | CCI | TCC | :ATC | TCG     | AAC | CTI | ccc | TCA | TG  | SACG | CCC | TGC | TGP  | ATC  | ACC  | CGG |
|      |      | 70   |     |     |     |      |         |     | 90  |     |     |     |      |     | 11  | 0    |      |      |     |
| CCIO | AG"  | TTCG | TGC | GCT | TGC | TCA  | TTT     | ccc | ACC | GCC | TCA | GCÇ | TGG  | GCC | ACI | TCC  | TGA  | CCC  | CG  |
|      |      | 130  |     |     |     |      |         | 1   | .50 |     |     |     |      |     | 17  | 0    |      |      |     |
| ATG  | GCC  | TGG  | ccc | AAC | TCT | ACA  | rece    | CGG | CGC | CCT | CCA | ACT | CGC  | TCA | TCC | :GCA | ACC  | TT   | TG  |
|      |      | 190  |     |     |     |      |         | 2   | 10  |     |     |     |      |     | 23  | 0    |      |      |     |
| GAC  | CAGO | CCT  | CCC | ACA | GCG | CAG  | GCA     | CCA | AAG | CCC | CAG | CCC | TAA  | AAG | GGG | GAG  | CTC  | CGG  | AG  |
|      |      | 250  |     |     |     |      |         | 2   | 70  |     |     |     |      |     | 29  | 0    |      |      |     |
| CTCC | GGC  | CCC  | CTG | ACG | TGG | GGC  | ATG     | TGC | TGA | GGA | TGC | TGC | TGG  | GGA | AGA | TGI  | 'GCG | CĢC  | ЖG  |
|      |      | 310  |     |     |     |      |         | 3   | 30  |     |     |     |      |     | 35  | 0    |      |      |     |
| AGAT | GT!  | TCT  | GCI | CTC | GGA | CAA  | GGC     | CAC | CTC | GCC | GCT | CTC | CT   | GGA | TGC | TGG  | CCI  | CG0  | GC  |
| M    | Y    | L    | L   | S   | D   | K    | A       | T   | S   | P   | L   | S   | L    | D   | A   | G    | L    | G    | Q   |
|      |      | 370  |     |     |     |      |         | 3   | 90  |     |     |     |      |     | 41  | 0    |      |      |     |
| AGGC | CCC  | CTG  | GAG | CGA | CCT | GCT  | TCT     | TTG | GGC | ACT | GTT | GCI | 'GAA | CAG | GGC | ACA  | GAT  | 'GGC | CA  |
| A    | P    | W    | s   | D   | L   | L    | ${f r}$ | M   | A   | L   | L   | L   | N    | R   | A   | Q    | M    | A    | М   |
|      |      | 430  |     |     |     |      |         | 4   | 50  |     |     |     |      |     | 47  | 0    |      |      |     |
| TGTA | CTI  | CTG  | GGA | GAT | GGG | TTC  | CAA     | TGC | AGT | TTC | CTC | AGC | TCT  | TGG | GGC | CTG  | TTT  | GCI  | 'GC |
| Y    | F    | W    | E   | М   | G   | S    | N       | A   | V   | S   | S   | A   | Ľ    | G   | A   | С    | L    | L    | L   |

Fig. 9 / continue 1 4

510 530 TCCGGGTGATGGCACGCCTGGAGCCTGACGCTGAGGAGGCACCGCAGGAAAGACCTGG RVMARLEPDAEEAARKDLA 590 570 CGTTCAAGTTTGAGGGGATGGGCGTTGACCTCTTTGGCGAGTGCTATCGCAGCAGTGAGG F K F E G M G V D L F G E C Y R S S E V 630 650 RAARLLLRRCPLWGDATCLQ 690 AGCTGGCCATGCAAGCTGACGCCCGTGCCTTCTTTGCCCAGGATGGGGTACAGTCTCTGC LAMQADARAFFAQDGVQSLL 750 770 TGACACAGAAGTGGTGGGGAGATATGGCCAGCACTACACCCATCTGGGCCCTGGTTCTCG TQKWWGDMASTTPIWALVLA 810 CCTTCTTTGCCCTCCACTCATCTACACCCGCCTCATCACCTTCAGGAAATCAGAAGAGG F F C P P L I Y T R L I T F R K S E E E 870 890 AGCCCACACGGGAGGAGCTAGAGTTTGACATGGATAGTGTCATTAATGGGGAAGGGCCTG PTREELEFDMDSVINGEGPV 930 950 TCGGGACGGGGCCAGCCGAGAAGACGCCGCTGGGGGTCCCGCGCCAGTCGGGCCGTC 990 1010 CGGGTTGCTGCGGGGGCCCGGTGCCTACGCCGCTGGTTCCACTTCT G C C G G R C G G R R C L R R W F H F W 1030 1050 1070  ${\tt GGGGGTGCCGGTGACCATCTTCATGGGCAACGTGGTCAGCTACCTGCTGTTCCTGCTGC}$ G V P V T I F M G N V V S Y L L F L L 1110 1130 F S R V L L V D F Q P A P P G S L E L L 1170 1190 TGCTCTATTTCTGGGCTTTCACGCTGCTGTGCGAGGAACTGCGCCAGGGCCTGAGCGGAG LYFWAFTLLCEELRQGLSGG 1230 1250 GCGGGGCAGCCTCGCCAGCGGGGCCCCGGGCCTGGCCATGCCTCACTGAGCCAGCGCC G G S L A 8 G G P G P G H A S L S Q R L 1270 1310 1290 TGCGCCTCTACCTCGCCGACAGCTGGAACCAGTGCGACCTAGTGGCTCTCACCTGCTTCC RLYLADSWNQCDLVALTCFL 1350 1370 TCCTGGGCGTGGGCTGCCGGCTGACCCCGGGTTTGTACCACCTGGGCCGCACTGTCCTCT LGVGCRLTPGLYHLGRTVLC 1410 1430 GCATCGACTTCATGGTTTTCACGGTGCGGCTGCTTCACATCTTCACGGTCAACAACAGC IDFMVFTVRLLHIFTVNKQL 1470 TGGGGCCCAAGATCGTCATCGTGAGCAAGATGATGAAGGACGTGTTCTTCCTCTTCT G P K I V I V S K M M K D V F F F L F F 1550 1530 TCCTCGGCGTGTGGCTGGCCTATGGCGTGGCCACGGAGGGGCTCCTGAGGCCACGGG LGVWLVAYGVATEGLLRPRD 1590 1610 ACAGTGACTTCCCAAGTATCCTGCGCCGCGTCTTCTACCGTCCCTACCTGCAGATCTTCG S D F P S I L R R V F Y R P Y L Q I F G 1630 1650 1670 GECAGATTCCCCAGGAGGACATGGACGTGGCCCTCATGGAGCACAGCAACTGCTCGTCGG Q I P Q E D M D V A L M E H S N C S S E 1730 1710 AGCCCGGCTTCTGGGCACACCCTCCTGGGGCCCAGGCGGGCACCTGCGTCTCCCAGTATG

WO 02/010382 PCT/EP01/08309

Fig. 9 / continuation 5 P G F W A H P P G A Q A G T C V S Q Y A 1790 1750 1770 CCAACTGGCTGGTGCTGCTCCTCGTCATCTTCCTGCTCGTGGCCAACATCCTGCTGG NWLVVLLLVIFLLVANILLV 1830 1850 TCAACTTGCTCATTGCCATGTTCAGTTACACATTCGGCAAAGTACAGGGCAACAGCGATC N L L I A M F S Y T F G K V Q G N S D L 1910 1890 TCTACTGGAAGGCGCAGCGTTACCGCCTCATCCGGGAATTCCACTCTCGGCCCGCGCTGG Y W K A Q R Y R L I R E F H S R P A L A 1970 1950 1930  $\tt CCCCGCCCTTTATCGTCATCTCCCACTTGCGCCTCCTGCTCAGGCAATTGTGCAGGCGAC$ P P F I V I S H L R L L L R Q L C R R P 2030 2010 1990 CCCGGAGCCCCAGCCGTCCTCCCCGGCCCTCGAGCATTTCCGGGTTTACCTTTCTAAGG R S P Q P S S P A L E.H F R V Y L S K E 2070 2090 2050  ${\tt AAGCCGAGCGGAAGCTGCTAACGTGGGAATCGGTGCATAAGGAGAACTTTCTGCTGGCAC}$ AERKLLTWESVHKENFLLA.R 2130 2150 GCGCTAGGGACAAGCGGAGAGCGACTCCGAGCGTCTGAAGCGCACGTCCCAGAAGGTGG ARDKRESDSERLKRTSQKVD 2210 2170 2190  ${\tt ACTTGGCACTGAAACAGCTGGGACACATCCGCGAGTACGAACAGCGCCTGAAAGTGCTGG}$ LALKQLGHIREYEQRLKVLE 2250 2270 REVQQCSRVLGWVAEALSRS 2330 2310 2290 ALLPPGGPPPPDLPGSKD\* 2390 2350 2370 CCCTGCTGGCGGACTTCAAGGAGAAGCCCCCACAGGGGATTTTGCTCCTAGAGTAAGGCT 2430 2450 CATCTGGGCCTCGGCCCCGCACCTGGTGGCCTTGTCCTTGAGGTGAGCCCCATGTCCAT 2490 2510 2470 CTGGGCCACTGTCAGGACCACCTTTGGGAGTGTCATCCTTACAAACCACAGCATGCCCGG 2570 2550 2530 CTCCTCCCAGAACCAGTCCCAGCCTGGGAGGATCAAGGCCTGGATCCCGGGCCGTTATCC 2630 2610 ATCTGGAGGCTGCAGGGTCCTTGGGGTAACAGGGACCACAGACCCCTCACCACTCACAGA 2690 2670 

MYLLSDKATSPLSLDAGLGQAPWSDLLLWALLINRAQMAMYFWEMGSNAVSSALGACLLLRVMARLEPDAEEAARRKDLAFKFEGM
GVDLFGECYRSSEVRAARLLLRRCPLWGDATCLQLAMQADARAFFAQDGVQSLLTQKWWGDMASTTPIWALVLAFFCPPLIYTRLI
TFRKSEEEPTREELEFDMDSVINGEGPVGTADPAEKTPLGVPRQSGRPGCCGGRCGGRRCLRRWFHFWGVPVTIFMGNVVSYLLFL
LLFSRVLLVDFQPAPPGSLELLLYFWAFTLLCEELRQGLSGGGGSLASGGPGPGHASLSQRLRLYLADSWNQCDLVALTCFLLGVG
CRLTPGLYHLGRTVLCIDFMVFTVRLLHIFTVNKQLGPKIVIVSKMMKDVFFFLFFLGVWLVAYGVATEGLLRPRDSDFPSILRRV
FYRPYLQIFGQIPQEDMDVALMEHSNCSSEPGFWAHPPGAQAGTCVSQYANWLVVLLLVIFLLVANILLVNLLIAMFSYTFGKVQG
NSDLYWKAQRYRLIREFHSRPALAPPFIVISHLRLLLRQLCRRPRSPQPSSPALEHFRVYLSKEAERKLLTWESVHKENFLLARAR
DKRESDSERLKRTSQKVDLALKQLGHIREYEQRLKVLEREVQQCSRVLGWVAEALSRSALLPPGGPPPPDLPGSKD

26/40

A)

|   |  | 10   |   |  |   |   |  | 30   |   |  |  |  |  | 50   |   |  |   |
|---|--|--|---|--|---|---|--|--|---|--|--|--|--|--|---|--|---|
| ATTA  | DAA  | TTT  | ATA   | AAA                                    | CAG   | TGG   | CTG  | GATGG  | TTGG  | AGG  | ATG  | CAG  | GTG  | GACAG  | AAGA  | CGT  | GG  |
|   |  |  |   |  |   |   |  | M V  | G   | G  | С  | R  | W  | T E  | D   | V  | E   |
| •   |  | 70   |   |  |   |   | ~  | 90   |   |  |  |  |  | 110  |   |  |   |
|   |  | :AGA<br>E  |   | 'AAA<br>K                              | GGA<br>E  | ааа<br>К  | GAT<br>M   | GTCCT<br>S F   |   |  |  | CAG<br>R   |  | CAGCA<br>S M   |   | GAA<br>N   | R   |
| F   | A  | 130  | ٧   | Д                                      | E   | I.  | 12   | 150  | Α.  | A  | A.   | K  |  | 170  |   | 14   | K   |
| GAAG  | GAZ  |  | CAC   | TCT                                    | GGA   | CAG   | CAC  |  | CCCI  | GTA  | CTC  | CAG  | CGC  | GTCTC  | GGAG  | CAC  | AG  |
|   |  | D  |   |  |   |   |  | R T  |   | Y  | _  |  |  |  |   | Т  | D   |
|   |  | 190  |   |  |   |   |  | 210  |   |  |  |  |  | 230  |   |  |   |
| ACTT  |  |  |   | TGA                                    | AAG   | CGC   |  |  |   | TGC  | CTT  | CAG  | GAC  | ACAGA  |   | CCC  | AA  |
| L   | S  | Y  | S   | E                                      | S   | A   | S  | F Y  | A   | A  | F  | R  | T  | -  | С   | P  | I   |
| TICATI  | cci  | 250  | mmc   | ייייי ז                                | cmm   |   | רא א   | 270  | mm/r)   | NCC  |  | mma  | መካክ  | 290<br>GAAAC   | C N C F   | אתר  | ·m~   |
| M   | oo.<br>A   |  | W   | D<br>D                                 | L   | V   |  | FI   |   | A  |  | F  |  |  | -   | C  | V.  |
| ••  | ••   | 310  | ••  | •                                      | -   | . •   | ••   | 330  | -   | ••   | •  | •  | **   | 350  |   | •  | •   |
| TCTT  | CT:  | CTAC   | CAP   | AGA                                    | TTC   | CAA   | GGC  | CACGG  | AGA?  | TGT  | 'GTG   | CAA  | GTG  | TGGCT  | ATGO  | CCA  | GA  |
| F   | F  | T  | K   | D                                      | S   | K   | A  | T E  | N   | V  | С  | K  | С  | G Y  | Α   | Q  | S   |
|   |  | 370  |   |  |   |   |  | 390  |   |  |  |  |  | 410  |   |  |   |
|   |  |  |   |  |   |   |  |  |   |  |  |  |  | CTACA  |   |  |   |
| Q   | н  | M<br>43D   | E   | G                                      | T   | Q   | Ţ  | N Q<br>450   |   | E  | ĸ  | W  | N  | Y K  | K   | Н  | T   |
| CCAA  | .GGZ   |  | TCC   | TAC                                    | CGA   | CGC   | стт  |  |   | тса  | GTT  | TGA  | GAC  | ACTGG  | GGAZ  | GAZ  | AG  |
| K   | E  |  | P   | T                                      | D   | A   |  | G D  |   | Q  | F  |  | T  | L G  |   | K  | G   |
|   |  | 490  |   |  |   |   |  | 510  |   |  |  |  |  | 530  |   |  |   |
| GGAA  | GT2  | TAT  | ACG   | TCT                                    | GTC   | CTG   |  |  |   | CGGA   | TAA  | CCI  | 'TTA   | .CGAGC   | TGCI  | GAC  | cc  |
| K   | Y  | _  | R   | L                                      | S   | С   | D  | T D  |   | E  | I  | L  | Y  | EL   | L   | T  | Q   |
| ACCA.   | cm/  | 550  |   | מ מיינו                                | B 7 C   | 200   | ~~~  | 570  |   | mme  | mem  | ~ * <i>~</i>   |  | 590<br>GGGCG   | ~~~ T   | ~ 7 7  | <u>~</u>  |
| H   |  | H  | L   | GAA<br>K                               | AAC<br>T  | P<br>P  |  | L V  |   |  | .1G1   |  |  |  |   | N  | F   |
|   |  | 610  | _   |  | _   | _   |  | 630  | _   | Ī  | •  | -  | •  | 650  |   |  | _   |
| TCGC  | CC:  | rga <b>a</b>   | GCC   | GCG                                    | CAT   | GCG   | CAA  | GATCT  | TCAC  | CCG  | GCT  | CAI  | CTA  | CATCG  | CGCF  | AGTO   | CA.   |
| A   | -  |  |   |  |   |   |  |  | _   | _  | •  | т  | Y  |  | _   | _  |   |
|   | L  | K  | P   | R                                      | M   | R   | K  | I F  | S   | R  | L  | _  | 1  |  | . Q   | S  | K   |
| አአሮሮ  |  | 670  |   |  |   |   | -  | 690  | _   | -  |  |  |  | 710  | _   |  |   |
|   | TGO  | 670<br>CTTG  | GAT   | TCT                                    | CAC   | GGG   | AGG  | 690<br>CACCC   | ATT!  | \TGG   | CCT  | GAT  | 'GAA   | 710<br>GTACA   | TCGC  |  |   |
| AAGG<br>G   | TG(  | 670<br>CTTG  | GAT   |  | CAC   | GGG   | AGG  | 690  | ATTI<br>Y   | -  | CCT  | GAT  |  | 710  | TCGC  | GG.  | ∆GG   |
| G<br>·  | TG(  | 670<br>TTG<br>W<br>730   | GAT<br>I  | TCT<br>L                               | CAC   | GGG<br>G  | AGG<br>G   | 690<br>CACCC<br>T H<br>750   | ATTI<br>Y   | ATGG<br>G  | CCT<br>L   | gat<br>M   | 'GAA<br>K  | 710<br>GTACA<br>Y I  | TCG(  | E<br>E   | V<br>VGG  |
| G<br>·  | TG(<br>A<br>GA(  | 670<br>TTG<br>W<br>730<br>AGA<br>D   | GAT<br>I<br>TAA   | TCT<br>L                               | CAC   | GGG<br>G  | AGG<br>G<br>CAG  | 690<br>CACCC<br>T H<br>750   | ATTI<br>Y<br>CAGI   | ATGG<br>G  | CCT<br>L   | gat<br>M   | 'GAA<br>K  | 710<br>GTACA<br>Y I<br>770   | TCGC<br>G                                       | E<br>E   | V<br>VGG  |
| g<br>TGGT<br>V  | TG(<br>A<br>GA(<br>R   | 670<br>TTG<br>W<br>730<br>SAGA<br>D<br>790   | gat<br>I<br>Taa<br>N  | TCT<br>L<br>CAC                        | CAC<br>T<br>CAT                                     | GGG<br>G<br>CAG<br>S  | AGG<br>G<br>CAG<br>R   | 690<br>CACCC<br>T H<br>750<br>GAGTT<br>S S   | ATTI<br>Y<br>CAGI<br>E  | ATGG<br>G<br>AGGA<br>E   | CCT<br>L<br>L<br>GAA   | TAD<br>M<br>TAT  | 'GAA<br>K<br>'TGT<br>V   | 710<br>GTACA<br>Y I<br>770<br>GGCCA<br>A I<br>830  | TCGG<br>G<br>TTGG                               | egg <i>e</i><br>E<br>ECAT  | AGG<br>V<br>PAG<br>A  |
| G<br>TGGT<br>V  | TGC<br>A<br>GAC<br>R   | 670<br>W<br>730<br>GAGA<br>D<br>790<br>GGGG  | GAT  TAA  N  CAT  | TCT<br>L<br>CAC<br>T                   | CAC T CAT   | GGG<br>G<br>CAG<br>S  | AGG<br>G<br>CAG<br>R   | 690 CACCC T H 750 GAGTT S S 810 GGACA  | ATTI<br>Y<br>CAGI<br>E  | ATGG<br>G<br>AGGA<br>E   | CAG  | CAT<br>M<br>TAT<br>I   | GAA<br>K<br>K<br>TGT<br>V  | 710<br>GTACA<br>Y I<br>770<br>GGCCA<br>A I<br>830<br>CGATG   | TCGG<br>G<br>TTGG<br>G                          | egga<br>E<br>ECAT<br>I   | AGG<br>V<br>PAG<br>A<br>A   |
| g<br>TGGT<br>V  | TG(<br>A<br>GA(<br>R   | 670<br>W<br>730<br>SAGA<br>D<br>790<br>GGGG  | GAT I TAA N CAT   | TCT<br>L<br>CAC                        | CAC<br>T<br>CAT                                     | GGG<br>G<br>CAG<br>S  | AGG<br>G<br>CAG<br>R   | 690 CACCC T H 750 GAGTT S S 810 GGACA D T  | ATTI<br>Y<br>CAGI<br>E<br>CCCI                                    | ATGG<br>G<br>AGGA<br>E   | CAG  | CAT<br>M<br>TAT<br>I   | 'GAA<br>K<br>'TGT<br>V   | 710 GTACA Y I 770 GGCCA A I 830 CGATG  | TCGG<br>G<br>TTGG<br>G                          | egg <i>e</i><br>E<br>ECAT  | AGG<br>V<br>PAG<br>A  |
| G<br>TGGT<br>V<br>CAGC                                | TGG<br>A<br>GAG<br>R<br>TTG<br>W   | 670<br>TTG<br>W<br>730<br>SAGA<br>D<br>790<br>GGGG<br>G  | GAT<br>I<br>TAA<br>N<br>CAT                                     | CAC<br>T<br>T<br>CGT<br>V              | CAC<br>T<br>CAT<br>I<br>CTC                         | GGG<br>G<br>CAG<br>S<br>CAA   | R<br>G<br>CAG<br>R<br>ACCG<br>R                                      | 690 CACCC T H 750 GAGTT S S 810 GGACA D T 870  | ATTI<br>Y<br>CAGI<br>E<br>CCCI                                    | ATGG<br>G<br>AGGA<br>E<br>CAT  | CCT<br>L<br>LGAA<br>N<br>CAG   | TAT<br>I<br>I<br>GAP   | GAA<br>K<br>TGT<br>V<br>TTG  | 710<br>GTACA<br>Y I<br>770<br>GGCCA<br>A I<br>830<br>CGATG<br>D A  | TCGG  | EGGA<br>E<br>ECAT<br>I<br>AGGG   | AGG<br>V<br>PAG<br>A<br>SCT<br>Y  |
| G<br>TGGT<br>V<br>CAGC                                | TGO<br>A<br>GAO<br>R<br>TTO<br>W   | 670<br>TTG<br>W<br>730<br>SAGA<br>D<br>790<br>GGGG<br>G<br>850<br>FAGC   | GAT TAM N CAT   | CAC<br>T<br>CGT<br>V                   | CAC<br>T<br>CAT<br>I<br>CTC<br>S                    | GGG<br>G<br>CAG<br>S<br>CAA<br>N                                      | AGG<br>G<br>CAG<br>R<br>ACCG<br>R                                    | 690 CACCC T H 750 GAGTT S S 810 GGACA D T 870  | ATTA Y CAGA E CCCT L  | ATGG<br>G<br>AGGA<br>E<br>CAT<br>I   | CCT<br>L<br>LGAA<br>N<br>CAG<br>R  | M TATI   | GAA<br>K<br>TGT<br>V<br>TTG<br>C   | 710<br>GTACA<br>Y I<br>770<br>GGCCA<br>A I<br>830<br>CGATG<br>D A<br>890   | TCCI  | EGGAT<br>EGGAT<br>I<br>AGGG  | AGG<br>V<br>PAG<br>A<br>SCT<br>Y  |
| G TGGT V CAGC A                                       | TGO<br>A<br>GAO<br>R<br>TTO<br>W   | 670<br>TTG<br>W<br>730<br>SAGA<br>D<br>790<br>GGGG<br>G<br>850<br>FAGC   | GAT I TAA N CAT M CCA   | CAC<br>T<br>CGT<br>V                   | CAC<br>T<br>CAT<br>I<br>CTC<br>S                    | GGG<br>G<br>CAG<br>S<br>CAA<br>N                                      | AGG<br>G<br>CAG<br>R<br>ACCG<br>R                                    | 690 CACCC T H 750 GAGTT S S 810 GGACA D T 870 TGACT  | ATTI Y CAGI E CCCT L TCAG   | ATGG<br>G<br>AGGA<br>E<br>CAT<br>I   | CCT<br>L<br>LGAA<br>N<br>CAG<br>R  | M TATI   | GAA<br>K<br>TGT<br>V<br>TTG<br>C   | 710<br>GTACA<br>Y I<br>770<br>GGCCA<br>A I<br>830<br>CGATG<br>D A<br>890   | TCCI  | EGGAT<br>EGGAT<br>I<br>AGGG  | AGG<br>V<br>PAG<br>A<br>SCT<br>Y  |
| G TGGT V CAGC A ATTT F                                | TGG<br>A<br>GAG<br>R<br>TTG<br>W   | 670<br>W 730<br>SAGA<br>D 790<br>SGGG<br>G 850<br>PAGC<br>ACACAC   | GAT I TAA N CAT M CCA   | CAC T CGGT V AGTA Y                    | CAC T CAT I CTC S CCT L                             | GGGG<br>G<br>CAG<br>S<br>CAA<br>N<br>TAT<br>M                         | AGG<br>G<br>CAG<br>R<br>CCG<br>R                                     | 690 CACCC T H 750 GAGTT S S 810 GGACA D T 870 TGACT D F 930 GGACA  | ATTIA Y CAGA E CCCCT L TCAC                                       | ATGG G AGGA E CAT I CAAG   | GCCT<br>L<br>LGAA<br>N<br>CAG<br>R<br>GAGA<br>D                            | M TATI   | GAA<br>K<br>TGT<br>V<br>TTG<br>C<br>C  | 710 GTACA Y I 770 GGCCA A I 830 CGATG D A 890 GTATA Y I 950  | TCGG  TTGG  CTGI  TCCT  L                       | E GCAI  LAGGG  G  CGA  TGGA  D   | AGG V PAG A A GCT Y ACA N   |
| G TGGT V CAGC A ATTT F                                | TGG<br>A<br>GAG<br>R<br>TTG<br>W   | 670 W 730 TTG W 730 T90 GGGG G 850 FAGC A 910 ACAC T   | GAT I TAA N CAT M CCA A CA H                                    | CAC T CGGT V AGTA Y                    | CAC T CAT I CTC S CCT L                             | GGGG<br>G<br>CAG<br>S<br>CAA<br>N<br>TAT<br>M                         | AGG<br>G<br>CAG<br>R<br>CCG<br>R                                     | 690 CACCC T H 750 GAGTT S S 810 GGACA D T 670 TGACT D F 930 GGACA D N  | ATTIA Y CAGA E CCCCT L TCAC                                       | ATGG G AGGA E CAT I CAAG   | GCCT<br>L<br>LGAA<br>N<br>CAG<br>R<br>GAGA<br>D                            | M TATI   | GAA<br>K<br>TGT<br>V<br>TTG<br>C<br>C  | 710 GTACA Y I 770 GGCCA A I 830 CGATG D A 890 GTATA Y I 950 TCCCA  | TCGG  TTGG  CTGI  TCCT  L                       | E GCAI  LAGGG  G  CGA  TGGA  D   | AGG V PAG A A GCT Y ACA N   |
| G TGGT V CAGC A ATTT F ACAA                           | TTC<br>R<br>TTC<br>W<br>TTT<br>L<br>CC2                                    | 6700 ZTTG W 7300 SAGA D 7900 GGGGG G 8500 PAGC A 9100 ACAC T 970   | GAT I TAA N CAT M CCA ACA                                       | CAC T GGT V GTA Y                      | CAC T CAT I CTC S CCT L                             | GGGG<br>G<br>CAG<br>S<br>CAA<br>N<br>TAT<br>M                         | AGG<br>G<br>CAG<br>R<br>CCG<br>R<br>GGA<br>D                         | 690 CACCC T H 750 GAGTT S S 810 GGACA D T 870 TGACT D F 930 GGACA D N  | ATTIN Y CAGN E CCCCT L TCAC T ATGG                                | G G AGGA E I CAN I | GCCT<br>L<br>LGAA<br>N<br>CAG<br>R<br>GAGA<br>D<br>TTCA                    | GAT  I  GAF  N  TCC  P  TCG  G   | CAACT L ACA  | 710 GTACA Y I 770 GGCCA A I 830 CGATG D A 890 GTATA Y I 950 TCCCA P T  | TCGC G G TTGCC G CTGH L CTGCV                   | GGGAT<br>I<br>AGGGG<br>G<br>TGGA<br>D  | AGG V  FAG A  GCT Y  ACA N  |
| G . TGGT V . CAGC A . ATTT F . ACAA . N . CAAA        | TTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA                                    | 6700 ZTTG W 7300 SAGA D 7900 GGGGG G 8500 PAGC A 9100 ACAC T 970   | GAT I TAA N CAT M CCA ACA H GAA                                 | CAC T GGTA Y TTTT L                    | CAC T CAT I CTC S CCT L GCT                         | GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG                                | AGGAAAGG   | 690 CACCC T H 750 GAGTT S S 810 GGACA D T 870 TGACT D F 930 GGACA D N 990 GTATA  | ATTIVY CAGA E CCCCT L TCAC T ATGG                                 | ATGG<br>G<br>AGGA<br>E<br>CAAG<br>R<br>CAAG<br>C   | GCCT<br>L<br>LGAA<br>N<br>CAG<br>R<br>AGAA<br>D<br>TCA                     | GAT  I  GAP  TCC  P  TCC  G  G  TCC  CAC   | GAAA K TGT V TTGT C CACT L GACA H  | 710 GTACA Y I 770 GGCCA A I 830 CGATG D A 890 GTATA Y I 950 TCCCA P T 1010 TCAAG                                   | TCGC G G TTGCC G CTGH L CTGV                    | GGGAT  GGGAT  GGGAT  GGGAT  CGGAT  CG | AGG V AGG A AGGT Y AGA N AGA A  |
| G . TGGT V . CAGC A . ATTT F . ACAA . N . CAAA        | TGGAGR TTTG W TTTT L CCA H GCT   | 6700 ZTTG W 7300 SAGA D 7900 GGGGG G 8500 PAGC A 9100 ACAC T 970   | GAT I TAA N CAT M CCA M CCA ACA H GAA N                         | CAC T GGTA Y TTTT L                    | CAC T CAT I CTC S CCT L GCT                         | GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG                                | AGGAAAGG   | 690 CACCC T H 750 GAGTT S S 810 GGACA D T 870 TGACT D F 930 GGACA D N 990 GTATA  | ATTIVY CAGINE E CCCTT L TCAG T ATGG G TCTG                        | ATGG<br>G<br>AGGA<br>E<br>CAAG<br>R<br>CAAG<br>C   | GCCT<br>L<br>LGAA<br>N<br>CAG<br>R<br>AGAA<br>D<br>TCA                     | GAT  I  GAP  TCC  P  TCC  G  G  TCC  CAC   | GAAA K TGT V TTGT C CACT L GACA H  | 710 GTACA Y I 770 GGCCA A I 830 CGATG D A 890 GTATA Y I 950 TCCCA P T 1010 CGTAGG                                  | TCGC G G TTGCC G CTGH L CTGV                    | GGGAT  GGGAT  GGGAT  GGGAT  CGGAT  CG | AGG V AGG A AGGT Y AGA N AGA A  |
| G . TGGT V . CAGC A . ATTT F . ACAA . N . CAAA . K    | TGG<br>A<br>GAG<br>R<br>TTG<br>W<br>TTG<br>L<br>CCA<br>H                   | 670<br>W 730<br>EAGA<br>D 790<br>GGGGGG<br>G 850<br>PACAC<br>T 970<br>PCCG<br>R  | GAT I TAA N CAT M CCA ACA H GAA                                 | CAC T GGTA Y TTTT L TCA Q              | CAC T CAT I CTC S CCT L GCT L                       | GGGG<br>G<br>CAG<br>S<br>CAA<br>N<br>TAT<br>M<br>GCT<br>L             | AGG<br>R<br>ACCG<br>R<br>CGGA<br>D<br>CGT<br>V                       | 690 CACCC T H 750 GAGTT S S 810 GGACA D T 870 TGACT D F 930 GGACA D N 990 GTATA Y I 1050   | ATTI  | ATGG G AGGA E CCAT I CAAAG R CCTGA   | GCCT<br>L<br>LGAA<br>N<br>CAG<br>R<br>AGA<br>D<br>.TCA<br>H<br>LGCG        | GAT  I  GAP  N  TCC  P  TCC  C  C  C  C  C  C  T   | CGAA<br>K<br>V<br>TGT<br>V<br>TTG<br>C<br>CACT<br>L<br>ACA<br>H  | 710 GTACA Y I 770 GGCCA A I 830 CGATG D A 890 GTATA Y I 950 TCCCA P T 1010 TCAAG                                   | TCGG G TTGG G CTGH L CTGT V                     | GGGAT  GGGAT  AGGG  G  TGGAT  CGAT  CCAT   | AGG V TAG A GCT Y ACA N AGG A   |
| G TGGT V CAGC A ATTT F ACAA N CAAA K ATGG             | TGG A GAG R TTG W TTTT L GCT H GCT   | 6700 TTG W 7300 TTG T7300 T900 T900 T900 TAGG B500 TAGC T 9700 TCCG R L0300  | GAT I TAA N CAT M CCA A CA A H GAA N                            | CAC T GGTA Y TTTT L TCA Q              | CAC T CAT I CTC S CCT L GCT L GCT L                 | CAGG CAGA N TAT M GCT L AGAA E  | AGGAA K  | 690 CACCC T H 750 GAGTT S S 810 GGACA D T 870 TGACT D F 930 GGACA D N 990 GTATA Y I 1050   | ATTIVY CAGA   | ATGG  G  E  CATG  CAAG  C  CATG  AGCTG  C  C  C  C  C  C  C  C  C  C  C  C  C  | GCCT<br>L<br>LGAA<br>N<br>CAG<br>R<br>GAGA<br>D<br>TTCA<br>H<br>LGCG       | GAT  M  TAT  I  GGAF  N  TCC  P  TGG  G  CAC  T  | GAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA   | 710 GTACA Y I 770 GGCCA A I 830 CGATG D A 890 GTATA Y I 950 TCCCA P T 1010 TCAAG Q D                               | TCGCCTGF  CTGF  CTGF  L  CTGF  V  ATTC          | GGGA<br>E<br>GGGA<br>GGGA<br>CGGA<br>CCCAA   | AGG V TAG A A CT Y AAG A A A A A A A A A A A A A A A A A  |
| G . TGGT V CAGC A ATTT F ACAA K ATGG G                | TGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG                                    | 6700 TTG W 7300 TTG T7900 PROGGGGG G RAGA P100 PROGGGG R P100 PROGGGGG R P100 PROGGGGGG R P100 PROGGGGGGG R P100 PROGGGGGGG R P100 PROGGGGGGG R P100 PROGGGGGGG R P100 PROGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG | GAT I TAM N CAT M CCAT ACA H GAA I GAA I GAA I                  | CACCC P                                | CAC T CAT I CTC S CCT L GCT L GCT L CAT I           | GGGG<br>G<br>CAG<br>S<br>CAA<br>N<br>TAT<br>M<br>GCT<br>L<br>AGA<br>E | RAGG<br>R<br>CCG<br>R<br>CCGGA<br>CGGAA<br>K                         | 690 CACCC T H 750 GAGTT S S 810 GGACA D T 870 TGACT D F 930 GGACA D N 991 TGACT TTTT TGACT TTTT TGACT TTTT TGACT TTTT TGACT TATT TGACT TTTT TGACT TTT TGACT | ATTIN Y CAGN E CCCCT T ATGG G S CCCC Q                            | G C C C C C C C C C C C C C C C C C C C  | GCCT<br>L<br>GAAA<br>N<br>CAGG<br>R<br>GAGAA<br>D<br>TCA<br>H<br>GCGG<br>R | GAT  I  GAP  TCC  P  TCC  TCC  TCC  TCC  TCC  TCC  | CGAA<br>K<br>TGT<br>V<br>TTG<br>C<br>CACT<br>L<br>ACA<br>H<br>TTAT<br>I  | 710 GTACA Y I 770 GGCCA A I 830 CGATG D A 890 GTATA Y I 950 TCCCA P T 1010 TCAAG Q D 1070 AGAGA E T                | TCGC G  TTGC G  CTGH  E  TCCT L  CTGT V  ATTC S | GGGAT  I AGGGG  G CGGAT  D CCGAT  N CGGAT  K   | AGG V  TAG A  GCT Y  ACA N  ACT Y  ACT Y  |
| G CCAT  | TGO A  | 6700 TTG W 7300 TTG T7900 FAGA D B500 FAGC A 9100 FACAC T 9700 FCCG R L0300 K L0900 ATAC   | GAT I TAM N CAT M CCA A CA A GAA H GAA I CTC                    | CACCC P                                | CAC T CAT I CTC S CCT L GCT L GCT L CAT I           | GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG                                | RAGG<br>G<br>CCAG<br>R<br>CCCG<br>R<br>CGGAA<br>K<br>CGTG<br>CCTTGAA | 690 CACCC T H 750 GAGTT S S 810 GGACA D T 870 TGACT D F 930 GGACA D N 991 TGACT TO TTTTT TO TTTTTT TTTTTTTTTTTTTTTTT   | ATTIM Y CAGM E CCCCT L TTCAC T ATTCCCC S CCCCM Q CCTTCC           | G C C C C C C C C C C C C C C C C C C C  | GCCT<br>L<br>LGAA<br>N<br>CCAG<br>R<br>AGCG<br>R<br>HAGG<br>G              | GAN ITAT I GAN N TCC P TGG G TGG G GGG T   | CGAACT C C CACT L LACA H TAT I KAAAA K   | 710 GTACA Y I 770 GGCCA A I 830 CGATG D A 890 GTATA Y I 950 TCCCA P T 1010 TCAAG Q D 1070 AGAGA E T 1130 AGGCT     | TCGCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG          | GGGAT<br>GGGAT<br>D<br>TCGAT<br>N<br>K   | AGG V  TAG A  GCT Y  ACA N  ACT Y  ACA A  ACT Y   |
| G G G G G G G G G G G G G G G G G G G                 | TGO<br>A<br>GAO<br>R<br>TTO<br>U<br>CCA<br>H<br>GCT<br>TGO<br>G<br>CA<br>N | 6700 TTG W 7300 FAGA D 7900 FAGG B 8500 FAGC T 9700 FCCG R 10300 FCCAA K 10900 FTAC T  | GAT I TAA N CAT M CCA M CCA A CA A CA T T T T T T T T T T T T T | CACCC P                                | CAC T CAT I CTC S CCT L GCT L GCT L CAT I           | GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG                                | RAGG<br>G<br>CCAG<br>R<br>CCCG<br>R<br>CGGAA<br>K<br>CGTG<br>CCTTGAA | 690 CACCC T H 750 GAGTT S S 810 GGACA D T GACT D F 930 GGACA Q N 990 GTATA Y I 1050 TTTTG F A 1110 AATTC   | ATTIN Y CAGN E CCCCT T ATGG G S CCCCA CCCCCCCCCCCCCCCCCCCCCCCCCCC | G C C C C C C C C C C C C C C C C C C C  | GCCT<br>L<br>LGAA<br>N<br>CCAG<br>R<br>AGCG<br>R<br>HAGG<br>G              | GAN ITAT I GAN N TCC P TGG G TGG G GGG T   | CGAACT C C CACT L LACA H TAT I KAAAA K   | 710 GTACA Y I 770 GGCCA A I 830 CGATG D A 890 GTATA Y I 950 TCCCA P T 1010 TCAAG Q D 1070 AGAGA E T 1130 AGGCT G S | TCGCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG          | GGGAT<br>GGGAT<br>D<br>TCGAT<br>N<br>K   | AGG V  TAG A  GCT Y  ACA N  ACT Y  ACA A  ACT Y   |
| G CCAT  | TGO A GAO R TTO W TTT L CCO H GCT TGO G CAM                                | 6700 TTG W 7300 FAGA D 7900 GGGGG R 9100 FAGA T 9700 FCCG R 10300 FCCAA K 10900 TT 1150  | GAT I TAM N CAT M CCA M CCA H GAA I GAT I CTC S                 | CAC T GGT V GTA Y TTCA Q CCCC P        | CAC T CAT I CTC S CCT L GCT L CAT I CAAT I CAAA     | GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG                                | RAGG<br>G<br>CCAG<br>R<br>CCCG<br>R<br>CGGAA<br>K<br>CGTG<br>C       | 690 CACCC T H 750 GAGTT S S 810 GGACA D T GACT D F 930 GGACA Q N 990 GTATA Y I 1050 TTTTG F A 1110 AATTC I P   | ATTIN Y CAGN E CCCCT T ATGG G TCTCT S CCCCX Q CTTGC               | G CANGE C C C C C C C C C C C C C C C C C C C  | CCCT<br>L<br>GAAAN<br>PCAG<br>R<br>AAGAA<br>H<br>AGCG<br>R<br>AAGG<br>V    | GAPATATI I GGAPATATI I TCC P TCC T TCC T TCC T TCC T TCC T TCC T T TCC T | CGAACT C C CACT L ACACT I CAACA H CAACA E  | 710 GTACA Y I 770 GGCCA A I 830 CGATG D A 890 GTATA Y I 1010 TCAAG Q D 1070 AGAGA E T 1130 AGGCT G S 1190          | TCGGGGGG  | GGGAAGGGAAGGGAAGGGAAGGGAAGGGAAGGGAAGGGAAGGGAAGGGAAGGGAAGGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGAAGGAAAGAAAGAAAGAAAGAAAA   | AGG V  TAG A  GCT Y  ACA N  ACT Y  AGG A  ACT Y  AGG A  ACT Y   |
| TGGT V CAGC A ATTT F ACAA N CAAA K ATGG G CCAT I TCGC | TGG A GAGGE TTGG W TTTT L CCG H GCT TGG CAM TGG TGG                        | 6700 TTG W 7300 FAGA D 7900 GGGG G 8500 FAGC T 9700 FCCG R 10300 FCCG R 10300 FCCG T 11500 FTAC T 11500  | GAT I TAM N CAT M CCM ACM H GAM I CTC S GAT                     | CAC T GGT V GTA Y TTCA Q CCCC P CCAT I | CAC T CAT I CTC S CCT L GCT L GCT L CAT I CAA K TAG | GGGG G CAG S CAA N TAT M GCT L AGA E TGT V AAAA N CCT                 | AGGA R CCGT V .GAAA K CGTG C .C                                      | 690 CACCC T H 750 GAGTT S S 810 GGACA D T 670 TGACT D F 930 GGACA D N 990 GTATA Y I 1050 TTTTG F A 1110 AATTC I P 1170 GGAGG   | ATTIN Y CAGN E CCCCT T ATGG G T CCCCX Q CTTGG                     | AGGA  CAAGGA  CAAGGA  CAAGGA  CAAGGA  COTGAAGGA  COTGAAGAAGAAGAAGAAGAAAAAAAAAAAAAAAAAAAAA   | CCT<br>L<br>GAAA<br>N<br>CAGA<br>R<br>AAGA<br>H<br>AGCG<br>R<br>AAGG<br>V  | GAT  I  GAP  N  TCC  P  TCC  T  TCC  T  TCC  T  TCC  T  TCC  T  T  | GAACA  TGT  V  TGT  C  C  CACT  L  CACACA  H  TAT  I  CAAAA  K  CGGAACA  CGCAACA  CGCAACA  CGCAACA  CGCAACA  CGCAACA  CGCAACA  CGCAACA  CGCAACA  CGCAACA  CGCACA  CGCAACA  CCCAACA  CCCCAACA  CCCAACA  CCCCAACA  CCCAACA  CCCCAACA  CCCAACA  CCCAACA  CCCAACA  CCCAACA | 710 GTACA Y I 770 GGCCA A I 830 CGATG D A 890 GTATA Y I 1010 TCAAG Q D 1070 AGAGA E T 1130 AGGCT G S 1190 ATCTT    | TCGGGGGGGCTGCC                                  | GGGAAGGGAAGGGAAGGGAAGGGAAGGGAAGGGAAGGGAAGGGAAGGGAAGGGAAGGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGAAGGAAAGAAAGAAAGAAAA   | AGG V AGG A AGG Y AGG A |
| G CCAT  | TGG A GAG R TTG W TTT L CCI H GCT TGG G TTGG TTGG TTGG TTGG TTGG D         | 6700 TTG W 7300 FAGA D 7900 GGGG G 8500 FAGC T 9700 FCCG R 10300 FCCG R 10300 FCCG T 11500 FTAC T 11500  | GAT I TAM N CAT M CCA ACA H GAA T I CTC S GAT I                 | CAC T GGT V GTA Y TTCA Q CCCC P CCAT I | CAC T CAT I CTC S CCT L GCT L GCT L CAT I CAA K TAG | GGGG G CAG S CAA N TAT M GCT L AGA E TGT V AAAA N CCT                 | AGGA R CCGT V .GAAA K CGTG C .C                                      | 690 CACCC T H 750 GAGTT S S 810 GGACA D T 670 TGACT D F 930 GGACA D N 990 GTATA Y I 1050 TTTTG F A 1110 AATTC I P 1170 GGAGG   | ATTIN Y CAGN E CCCCT T ATGG G CCCCC Q CTTGG E                     | AGGA  CAAGGA  CAAGGA  CAAGGA  CAAGGA  COTGAAGGA  COTGAAGAAGAAGAAGAAGAAAAAAAAAAAAAAAAAAAAA   | CCT<br>L<br>GAAA<br>N<br>CAGA<br>R<br>AAGA<br>H<br>AGCG<br>R<br>AAGG<br>V  | GAT  I  GAP  N  TCC  P  TCC  T  TCC  T  TCC  T  TCC  T  TCC  T  T  | GAACA  TGT  V  TGT  C  C  CACT  L  CACACA  H  TAT  I  CAAAA  K  CGGAACA  CGCAACA  CGCAACA  CGCAACA  CGCAACA  CGCAACA  CGCAACA  CGCAACA  CGCAACA  CGCAACA  CGCACA  CGCAACA  CCCAACA  CCCCAACA  CCCAACA  CCCCAACA  CCCAACA  CCCCAACA  CCCAACA  CCCAACA  CCCAACA  CCCAACA | 710 GTACA Y I 770 GGCCA A I 830 CGATG D A 890 GTATA Y I 1010 TCAAG Q D 1070 AGAGA E T 1130 AGGCT G S 1190          | TCGGGGGGGCTGCC                                  | GGGAAGGGAAGGGAAGGGAAGGGAAGGGAAGGGAAGGGAAGGGAAGGGAAGGGAAGGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGAAGGAAAGAAAGAAAGAAAA   | AGG V AGG A AGG Y AGG A |

Fig. 10 / continue an 1

 ${\tt AGGAGAAGCTGGTGCGCTTTTTACCCCGGCAGGGTGTCCCGGCTGCCTGAGGAGGAGGAGCTG}$ EKLVRFLPRTVSRLPEETE 1290 1310 AGAGTTGGATCAAATGGCTCAAAGAAATTCTCGAATGTTCTCACCTATTAACAGTTATTA SWIKWLKEILECSHLLTVIK 1350 1370 AAATGGAAGAAGCTGGGGATGAAATTGTGAGCAATGCCATCTCCTACGCTCTATACAAAG MEEAGDEIVSNAISYALYKA 1410 1430 CCTTCAGCACCAGTGAGCAAGACAAGGATAACTGGAATGGGCAGCTGAAGCTTCTGCTGG F S T S E Q D K D N W N G Q L K L L E 1450 1470 1490 AGTGGAACCAGCTGGACTTAGCCAATGATGAGATTTTCACCAATGACCGCCGATGGGAGA WNQLDLANDEIFTNDRRWEK 1530 1550 AGAGCAAACCGAGGCTCAGAGACACAATAATCCAGGTCACATGGCTGGAAAATGGTAGAA S K P R L R D T I I Q V T W L E N G R I 1590 1610 TCAAGGTTGAGAGCAAAGATGTGACTGACGGCAAAGCCTCTTCTCATATGCTGGTGGTTC K V E S K D V T D G K A S S H M L V V L 1650 1670 TCAAGTCTGCTGACCTTCAAGAAGTCATGTTTACGGCTCTCATAAAGGACAGACCCAAGT K S A D L Q E V M F T A L I K D R P K F 1710 · 1730 TTGTCCGCCTCTTCTGGAGAATGGCTTGAACCTACGGAAGTTTCTCACCCATGATGTCC V R L F L E N G L N L R K F L T H D V L 1750 1770 1790 TCACTGAACTCTTCTCCAACCACTTCAGCACGCTTGTGTACCGGAATCTGCAGATCGCCA TELFSNHFSTLVYRNLQIAK 1830 1850 AGAATTCCTATAATGATGCCCTCCTCACGTTTGTCTGGAAACTGGTTGCGAACTTCCGAA NSYNDALLTFVWKLVANFRR 1870 1890 1910 GAGGCTTCCGGAAGGAAGACAGAAATGGCCGGGACGAGATGGACATAGAACTCCACGACG G F R K E D R N G R D E M D I E L H D V 1950 TGTCTCCTATTACTCGGCACCCCCTGCAAGCTCTCTTCATCTGGGCCATTCTTCAGAATA S P I T R H P L Q A L F I W A I L Q N K 1990 2010 2030 AGAAGGAACTCTCCAAAGTCATTTGGGAGCAGACCAGGGGCTGCACTCTGGCAGCCCTGG KELSKVIWEQTRGCTLAALG 2050 2070 2090 GAGCCAGCAAGCTTCTGAAGACTCTGGCCAAAGTGAAGAACGACATCAATGCTGCTGGGG A S K L L K T L A K V K N D I N A A G E 2130 2150 AGTCCGAGGAGCTGCTAATGAGTACGAGACCCGGGCTGTTGGTGAGTCCACAGTGTGGA S E E L A N E Y E T R A V G E S T V W N 2190 2210 ATGCTGTGGTGGCGCGGATCTGCCATGTGGCACAGACATTGCCAGCGGCACTCATAGAC AVVGADLPCGTDIASGTHRP 2250 2270 CAGATGGTGGAGAGCTGTTCACTGAGTGTTACAGCAGCGATGAAGACTTGGCAGAACAGC DGGELFTECYSSDEDLAEQL 2310 TGCTGGTCTATTCCTGTGAAGCTTGGGGTGGAAGCAACTGTCTGGAGCTGGCGGTGGAGG LVYSCEAWGGSNCLELAVEA 2370 2390  ${\tt CCACAGACCAGCATTCATCGCCCAGCCTGGGGTCCAGAATTTTCTTAAGCAATGGT}$ T D Q H F I A Q P G V Q N F L S K Q W Y 2410 2430 2450 ATGGAGAGATTTCCCGAGACACCAAGAACTGGAAGATTATCCTGTGTCTGTTTATTATAC GEISRDTKN W KILLCLFII P

Fig. 10 / continuation 2

|                              | 2470  |  |                                      |  |                                  |   | 2490  |   |                                       |   |   |  | 2510  |  |   |  |                                     |
|------------------------------|---|--|--------------------------------------|--|----------------------------------|---|---|---|---------------------------------------|---|---|--|---|--|---|--|-------------------------------------|
| CCTT                         | GGTGGG  | CTG  | TGG                                  | CTT                                      | rgt.                             | ATC   | ATTTAG  | GAA   | GAA                                   | ACC.  | rgt   | CGA  | CAAG  | CAC  | AA:   | AA   | 3C                                  |
| L                            |   | С  | G                                    | F  | V                                | S   | F R<br>2550   | ĸ   | ĸ                                     | P   | v   | Ð  | К 1<br>2570   | H  | K   | K  | L                                   |
| mcom                         | 2530  | <b>ረ</b> መአ  | mem.                                 |  |                                  | ~~~   |   | ~~~   | COMO.                                 | 000   | ~~m   | cour   |   |  | יתתי  | e C ma   | -0                                  |
|                              | TTGGTA  |  |                                      |  |                                  |   |   |   |                                       |   |   |  |   |  |   |  | ىرى<br>V                            |
| L                            | ₩ Y<br>2590   |  | ٧                                    | A  | F.                               | ť   | T S<br>2610   | r   | ¥                                     | V   | ٧   | F.   | S 1<br>2630   | N  | N   | V  | V                                   |
| TCTT                         | CTACAT  | CGC  | CTT                                  | CCT                                      | CCT                              | GCT   | STTTGC  | CTA   | CGT                                   | GCT   | GCT   | CAT  | GGAT'   | TTC  | CA'   | OTI  | GG                                  |
| F                            | ΥI  | Α  | F                                    | L  | L                                | L   | F A   | Y   | v                                     | L   | L   | М  | D :   | F  | H   | s  | v                                   |
| -                            | 2650  |  |                                      |  |                                  |   | 2670  |   |                                       |   | -   |  | 2690  |  |   |  |                                     |
| TGCC                         | ACACCC  | ccc  | CGA                                  | GCT                                      | GGT                              | CCT   |   | GCT   | CCT                                   | CTT   | rgr   | cci  | CTTC  | TG   | 'GA'  | rga.   | AG                                  |
| Þ                            | Н Р   |  |                                      |  |                                  |   | y s   | L   |                                       |   |   |  | F   |  | D   | E  | V                                   |
| •                            | 2710  | -  | _                                    |  | •                                | _   | 2730  | -   | •                                     | _   | •   | ~  | 2750  | •  | _   | _  | •                                   |
| CO CO                        | ACAGGG  |  |                                      |  | mr.c                             | er-C-C  |   |   |                                       | ~~  | ሮአክ   | ccc  |   | ~~   | יארי  | ccc  | CD.                                 |
|                              |   |  |                                      |  | -                                |   | -   |   | -                                     |   |   |  |   | CC<br>P  | T   | R  | N                                   |
| R                            | Q G   | R  | P                                    | A  | A                                | ٤   | S A   | G   | P                                     | A   | V   | ۲  |   | P  | 1   | K  | 14                                  |
|                              | 2770  |  |                                      |  |                                  |   | 2790  |   |                                       |   |   |  | 2810  |  |   |  |                                     |
|                              | CCATCTG   |  |                                      |  |                                  |   |   |   |                                       |   |   |  |   |  |   |  |                                     |
| S                            | I W   | _  | A                                    | ន  | S                                | T   | R S   | P   | G                                     | S   | R   | S  | R   |  | S   | F  | H                                   |
|                              | 2830  |  |                                      | •  |                                  |   | 2850  |   |                                       |   |   |  | 2870  |  |   | ٠  |                                     |
| ACAC                         | CTTCCCT   | GCA  | AGC                                  | TGA                                      | GGG                              | TGC   | CAGCTC  | TGG   | CCT                                   | TGG   | CCA   | GCC  | CAGA  | AA   | GG  | GTG  | GA                                  |
| Ŧ                            | S L   | Q  | A                                    | E  | G                                | A   | 8 8   | G   | L                                     | G   | Q   | P  | Ŗ   | K  | G   | W  | T                                   |
|                              | 2890  | )  |                                      |  |                                  |   | 2910  |   |                                       |   |   |  | 2930  |  |   |  |                                     |
| CAT                          | aaaaati   | TCI  | GGA                                  | AAT                                      | GGT                              | TGA   | TATTTC  | CAA   | GCT                                   | CCT   | GAT   | GI   | CCTC  | TC   | rgr   | CCC  | TT                                  |
| F                            | K N   | L  | E                                    | M  | V                                | D   | I S   | K   | L                                     | L.  | M   | S  | L   | S  | V   | P  | F                                   |
|                              | 2950  | )  |                                      |  |                                  |   | 2970  |   |                                       |   |   |  | 2990  |  |   |  |                                     |
| TCTC                         | STACGCA   | GTG  | GTA                                  | CGT                                      | AAA                              | TCC   | GGTGAA  | TTA   | TTT                                   | TAC   | TGA   | CC?  | rgt gg  | AA'  | rgt   | GAT  | GG                                  |
| С                            | T Q   | W  | Y                                    | v  | N                                | G   | и и   | Y   | F                                     | T   | D   | L  | M   | N  | V   | M  | D                                   |
|                              | 3010  | •  |                                      |  |                                  |   | 3030  |   |                                       |   |   |  | 3050  | i  |   |  |                                     |
| ACAC                         | CGCTGGG   | GCI  | ттт                                  | TTA                                      | CTI                              | CAT   | AGCAGG  | TAA   | TGT                                   | ATT   | TCG   | GC2  | AAGGG   | TA   | ccr   | TAC  | GC                                  |
| T                            | L G   | L  | F                                    | Y  | F                                | 1   | A G   | I   | v                                     | F   | R   | 0  | G   | I  | L   | R  | 0                                   |
|                              | 3070  |  | _                                    |  |                                  | -   | 3090  | _   |                                       |   |   | -  | 3110  |  |   |  | _                                   |
| AGAI                         | ATGAGCA   |  | CTG                                  | GAG                                      | GTG                              | GAT   | •   | TTC   | GGT                                   | САТ   | CTA   | CGI  |   |  | CCT   | GGC  | CA                                  |
| N                            | E Q   |  | W                                    | R  | W                                | I   | FR  | s   | v                                     |   | Y   |  |   |  | L   | A  | M                                   |
|                              | 3130  |  | ••                                   |  | ••                               | _   | 3150  | _   | ٠                                     | _   | -   | _  | 3170  | -  | _   |  |                                     |
| di Cale                      | rcggcca   |  | rece.                                | יר אכ                                    | TO                               | ССТ   |   | <u>ም</u> ልሰ   | יראכ                                  | ርሞል   | ጥርው   | البات  |   |  | ርጥር   | CAC  | CT.                                 |
| F                            |   | v.   |                                      | S  |                                  |   | D G   |   | T                                     | Y   |   |  | A   |  | c   |  | F                                   |
| •                            | 3190  |  | •                                    | J  | -                                | •   | 3210  | •   | •                                     | •   | _   | -  | 3230  |  | •   | •  | -                                   |
| mcn/                         | CTGGGAA   |  | CEC                                  | יראי                                     | ccc                              | ·» ~~   |   | PCC N   | ~~m                                   | VC/C3N  | מים   | cci  |   |  | 200   | ccc  | C.O.                                |
|                              |   |  |                                      |  |                                  |   |   |   |                                       |   |   |  |   | L  | P<br>P  | R  | F                                   |
| Ť                            | G N   |  | s                                    | K  | P                                | L   | C V<br>3270   | E   | L                                     | n   | E   | н  |   | _  |   | ĸ  | E                                   |
|                              | 3250  |  |                                      |  |                                  |   |   |   |                                       |   |   |  |   |  | E   |  |                                     |
|                              | CCGAGTG   |  |                                      |  |                                  |   |   |   |                                       |   |   |  | 3290  |  |   |  |                                     |
| P                            |   |  |                                      |  |                                  | _   | GGTGT   |   |                                       |   |   |  | CACC  | AA   | CAT   |  |                                     |
|                              | R W   | 1  | CAO<br>T                             | CAT<br>I                                 | P<br>CCC                         | _   | GGTGTC<br>V C   |   |                                       |   | GTI<br>L  |  | CACC  | AA<br>N  |   |  | EC.                                 |
|                              | 3310  | ı<br>)   | T                                    | I  | P                                | L   | GGTGTC<br>V C<br>3330   | I   | Y                                     | M   | L   | s  | CACC<br>T<br>3350   | AA<br>N  | CAT<br>I  | L  | L                                   |
| TGG:                         |   | ı<br>)   | T                                    | I  | P                                | L   | GGTGTC<br>V C<br>3330   | I   | Y                                     | M   | L   | s  | CACC<br>T<br>3350   | AA<br>N  | CAT<br>I  | L  | L                                   |
|                              | 3310  | I<br>GC1   | T<br>rgg1                            | I  | P<br>CAT                         | L<br>GPT  | GGTGTG<br>V C<br>3330<br>TGGCTI   | I   | Y<br>XGT                              | M<br>GGG  | L<br>CAC  | s<br>ccs   | T<br>3350<br>FCCAG  | aai<br>N<br>Gai  | CAT<br>I<br>GAA   | L<br>CAF   | L<br>TG                             |
|                              | 3310<br>CAACCI  | I<br>GC1   | T<br>rgg1                            | I  | P<br>CAT                         | L<br>GPT  | GGTGTG<br>V C<br>3330<br>TGGCTI   | I   | Y<br>XGT                              | M<br>GGG  | L<br>CAC  | s<br>ccs   | T<br>3350<br>FCCAG  | :AAI<br>N<br>GAI<br>E  | CAT<br>I<br>GAA   | L<br>CAF   | L<br>TG                             |
| v                            | 3310<br>CAACCT<br>N L   | I<br>GCT<br>L  | T<br>TGGT<br>V                       | I<br>CGC<br>A                            | P<br>CAI<br>M                    | L<br>GPT<br>F   | GGTGTG V C 3330 TGGCTI G Y 3390   | I<br>ACAC<br>T  | Y<br>XSGT<br>V                        | M<br>GGG<br>G                                       | L<br>CAC<br>T   | s<br>cg:<br>v  | T<br>3350<br>FCCAG<br>Q<br>3410   | AAI<br>N<br>GAI<br>E   | CAT<br>I<br>SAA<br>N  | L<br>Caf<br>N                                    | L<br>TG<br>D                        |
| V                            | 3310<br>CAACCT<br>N L<br>3370   | I<br>GCT<br>L<br>)<br>GGA  | T<br>TGGT<br>V<br>AGTT               | I<br>CGC<br>A                            | P<br>CAI<br>M                    | L<br>GTT<br>F<br>GTA  | GGTGTG V C 3330 TGGCTI G Y 3390   | I<br>ACAC<br>T  | Y<br>XGGT<br>V                        | M<br>GGG<br>G<br>.GGA                               | L<br>CAC<br>T   | S<br>CG:<br>V  | CCACC T 3350 FCCAG Q 3410 GCAGC   | AAI<br>N<br>GAI<br>E   | Cat<br>I<br>Gaa<br>N<br>CCT   | L<br>CAF<br>N                                    | L<br>TG<br>D                        |
| V                            | 3310<br>ICAACCI<br>N L<br>3370<br>AGGTCTO   | I<br>GCT<br>L<br>)<br>GGA!<br>K  | T<br>TGGT<br>V<br>AGTT               | I<br>CGC<br>A                            | P<br>CAI<br>M                    | L<br>GTT<br>F<br>GTA  | GGTGTG V C 3330 TGGCTI G Y 3390   | I<br>ACAC<br>T  | Y<br>XGGT<br>V                        | M<br>GGG<br>G<br>.GGA                               | L<br>CAC<br>T   | S<br>CG:<br>V  | CCACC T 3350 FCCAG Q 3410 GCAGC   | AAI<br>N<br>GAI<br>E<br>CGI                                  | Cat<br>I<br>Gaa<br>N<br>CCT   | L<br>CAF<br>N                                    | L<br>TG<br>D                        |
| V<br>ACCA<br>Q               | 3310<br>FCAACCT<br>N L<br>3370<br>AGGTCTG<br>V W<br>3430                                    | I<br>GCT<br>L<br>GGAA<br>K   | T<br>TGGT<br>V<br>AGTT<br>F          | I<br>CGC<br>A<br>CCA                     | P<br>CAT<br>M<br>GAG             | L<br>GPT<br>F<br>GTA<br>Y   | GGTGTG V C 3330 TGGCTI G Y 3390 CTTCCI F L 3450   | i<br>ACAC<br>T<br>TGGI<br>V                                       | Y<br>XGGT<br>V<br>XGCA<br>Q           | M<br>GGG<br>G<br>.GGA                               | L<br>CAC<br>T<br>GTA  | S<br>V<br>LCT(                                       | T<br>3350<br>FCCAG<br>Q<br>3410<br>SCAGO<br>S   | AAI<br>N<br>GAI<br>E<br>CGI<br>R                             | CAT<br>I<br>SAA<br>N<br>CCT<br>L  | CAF<br>N<br>CAF                                  | L<br>TG<br>D                        |
| V<br>ACCA<br>Q<br>TCC        | 3310<br>ICAACCI<br>N L<br>3370<br>AGGTCTG<br>V W<br>3430                                    | I<br>GCT<br>L<br>GGA!<br>K   | T<br>IGGI<br>V<br>AGTI<br>F          | I<br>CGC<br>A<br>CCCA<br>Q               | P<br>CAI<br>M<br>GAC<br>R        | L<br>F<br>GTA<br>Y  | GGTGTG V C 3330 TGGCTF G Y 3390 CTTCCF F L 3450 TTACTT  | I<br>ACAC<br>T<br>IGGI<br>V                                       | Y<br>XSGT<br>V<br>CGCA<br>Q           | M<br>GGG<br>G<br>.GGA<br>E                          | L<br>CAC<br>T<br>GTA<br>Y<br>GGT  | S<br>V<br>LCT(<br>C                                  | CCACC T 3350 CCAG Q 3410 SCAGC S 3470 AGAAG   | CAAI   | CAT I GAA N CCT   | CAF<br>N<br>CAF<br>N                             | L<br>TG<br>D<br>TA<br>I             |
| V<br>ACCA<br>Q<br>TCC        | 3310<br>FCAACCT<br>N L<br>3370<br>AGGTCTG<br>V W<br>3430                                    | I<br>GCT<br>L<br>GGA!<br>K   | T<br>IGGI<br>V<br>AGTI<br>F          | I<br>CGC<br>A<br>CCCA<br>Q               | P<br>CAI<br>M<br>GAC<br>R        | L<br>F<br>GTA<br>Y  | GGTGTC V C 3330 TGGCTI G Y 3390 CTTCCI F L 3450 TTACTI Y F  | I<br>ACAC<br>T<br>IGGI<br>V                                       | Y<br>XSGT<br>V<br>CGCA<br>Q           | M<br>GGG<br>G<br>.GGA<br>E                          | L<br>CAC<br>T<br>GTA<br>Y<br>GGT  | S<br>V<br>LCT(<br>C                                  | CCACC T 3350 CCAG Q 3410 SCAGC S 3470 AGAAG   | CAA  | CAT I GAA N CCT   | CAF<br>N<br>CAF<br>N                             | L<br>TG<br>D<br>TA<br>I             |
| V ACCI Q TCCC                | 3310<br>FCAACCT<br>N L<br>3370<br>AGGTCTO<br>V W<br>3430<br>CCTTCCO<br>F P<br>3490          | I<br>GCT<br>L<br>SGAM<br>K<br>)<br>CCTT  | T<br>V<br>AGTT<br>F<br>FCAT          | I<br>CCCA<br>CCCA<br>Q<br>CCCI           | P<br>CAT<br>M<br>GAG<br>R<br>CTT | L<br>F<br>F<br>GGTA<br>Y<br>CGC<br>A                              | GGTGTC<br>V C<br>3330<br>TGGCTI<br>G Y<br>3390<br>CTTCCI<br>F L<br>3450<br>TTACTI<br>Y F<br>3510                  | I<br>ACAC<br>T<br>FGGT<br>V<br>FCTA                               | Y<br>V<br>GCA<br>Q<br>CAT<br>M        | M<br>GGG<br>G<br>LGGA<br>E<br>'GGT'<br>V            | L<br>CAC<br>T<br>GTA<br>Y<br>GGT  | S<br>V<br>LCTC<br>C<br>C<br>GAI                      | 3350<br>FCCAG<br>Q<br>3410<br>GCAGC<br>S<br>3470<br>AGAAG<br>K<br>3530                                      | CAA  | CAT I GAA N CCT L CTT   | CAF<br>N<br>CAF<br>N<br>CAF<br>K                 | L<br>TG<br>D<br>TA<br>I<br>CT       |
| V ACCI Q TCCC P              | 3310 PCAACCT N L 3370 AGGTCTG V W 3430 CCTTCCC F P 3490 GCTGCAA                             | I<br>FGCT<br>K<br>K<br>CCTT<br>F   | T<br>V<br>AGTT<br>F<br>FCAT<br>I     | I<br>CCGC<br>A<br>CCCA<br>Q<br>CCGT<br>V | P CAT R CTT F                    | EGTA Y CGC A  | GGTGTG V C 3330 TGGCTF G Y 3390 CTTCCT F L 3450 TTACTT Y F 3510 GTCTTC  | I<br>ACAC<br>T<br>TGGI<br>V<br>ICTA<br>Y                          | Y XGCA Q ACAT M                       | M<br>GGG<br>G<br>GGA<br>E<br>GGT<br>V               | L CAC   | S<br>V<br>LCTO<br>C<br>C<br>TGAL                     | CCACO T 3350 CCAG Q 3410 GCAG S 3470 AGAAG K 3530 GGTTT   | AAN CAA CCG CCG CTG CCG                                      | CAT I GAA N CCT F   | L<br>CAF<br>N<br>CAF<br>N<br>CAF<br>K            | L<br>TG<br>D<br>TA<br>I<br>CT<br>C  |
| V ACCI Q TCCC                | 331C N L 337C AGGTCTG V W 343C CCCTTCCC F P 349C GCTGCAP C K                                | I CGCT K CCTT F AGGA   | T<br>V<br>AGTT<br>F<br>FCAT<br>I     | I<br>CCGC<br>A<br>CCCA<br>Q<br>CCGT<br>V | P CAT R CTT F                    | EGTA Y CGC A  | GGTGTG V C 3330 TGGCTF G Y 3390 CTTCCF F L 3450 TTACTF Y F 3510 GTCTTC S S  | I ACACAC T TGGI V TCTA  | Y XGCA Q ACAT M                       | M<br>GGG<br>G<br>GGA<br>E<br>GGT<br>V               | L CAC   | S<br>V<br>LCTO<br>C<br>C<br>TGAL                     | CCACO T 3350 CCAG Q 3410 CCAG S 3470 AGAAG K 3530 GGTTT F   | RAAI<br>CGA<br>CGG<br>R<br>CGG<br>C                          | CAT I GAA N CCT F   | L<br>CAF<br>N<br>CAF<br>N<br>CAF<br>K            | L<br>TG<br>D<br>TA<br>I<br>CT<br>C  |
| V ACCI Q TCCC P GTTC         | 331C TCAACCT N L 337C AGGTCTC V W 3430 CCTTCCC F P 3490 SCIGCAP C K 3550                    | I  GGCT  K  CCCTT  F  AGGS  E  | T (GGT V AGTT F (CAT I AGAA          | I CCCA A CCCA Q CCCT V                   | P CAT F CAT                      | L GGTA Y CGGC A CGGA E  | GGTGTC V C 3330 TGGCTH G Y 3390 CTTCCT F L 3450 TTACTT Y F 3510 GTCTTC S S 3570                                   | I ACACAC T TGGT V TCTA Y TGGT                                     | Y CGCA Q CAT M CTG                    | GGGA<br>GGGA<br>E<br>GGTG<br>V                      | L CAC   | S<br>CCGC<br>V<br>CCTC<br>C<br>CCAA<br>K<br>GTC<br>W | 2350<br>FCCAG<br>Q<br>3410<br>SCAGO<br>S<br>3470<br>AGAAG<br>K<br>3530<br>F<br>GGTTT<br>F                   | RAAI<br>N CGA<br>E CCG<br>R C C                              | CAT I SAA N CCT L CTT F CCA   | CAF<br>N<br>CAF<br>N<br>CAF<br>K                 | L<br>TG<br>D<br>ATA<br>I<br>C<br>C  |
| V ACCA  Q TCCC  P GTTC  C    | 331C TCAACCT N L 337C AGGTCTC V W 3430 CCCTTCCC F P 3490 SCIGCAP C K 3550                   | I  CGCT  K  CCCTT  F  AGGG  CAGGG  CA | T TGGTT V AGTT F TCAT I AGAA K AAGG  | I CCGC A CCGA Q CCGT V AAAA N            | P CAT R CTT F CAT M              | L COTT F F GGTA Y CGC A CGGA E TAA                                | GGTGTC V C 3330 TGGCTH G Y 3390 CTTCCT F L 3450 TTACTT Y F 3510 GTCTTC S S 3570 TTTCAC                            | I ACACAC T T T GGTI V T CTAI Y CTGTI V                            | Y CGCA Q LCAT M CTGC                  | M GGGG G GGA E GGT V CTG C ATG                      | L CAC   | S CCGS V CCTC C C C K GTC W                          | 2350<br>FCCAG<br>Q<br>3410<br>SCAGO<br>S<br>3470<br>AGAAG<br>K<br>3530<br>SGTTT<br>F<br>3590<br>ATCCA       | RAAI<br>N E C C R C C AT I                                   | CAT<br>I<br>GAA<br>N<br>CCT<br>L<br>CTT<br>F<br>CCA<br>H  | CAF<br>N<br>CAF<br>N<br>CAF<br>K<br>TGT<br>V     | L ATG D ATA I CT C C CT Y           |
| V ACCI Q TCCC P GTTC         | 331C TCAACCT N L 337C AGGTCTCC V W 3430 CCTTCCC F P 3490 GCTGCAA C K 3550 TGGGGATC G S      | I  CGCT  K  CCCTT  F  AGGI  E  CAGI  | T TGGTT V AGTT F TCAT I AGAA K AAGG  | I CCGC A CCGA Q CCGT V AAAA N            | P CAT R CTT F CAT M              | L COTT F F GGTA Y CGC A CGGA E TAA                                | GGTGTC V C 3330 TGGCTX G Y 3390 CTTCCI F L 3450 TTACTI Y F 3510 GTCTTC S S 3570 TTTCAG F R                        | I ACACAC T T T GGTI V T CTAI Y CTGTI V                            | Y CGCA Q LCAT M CTGC                  | M GGGG G GGA E GGT V CTG C ATG                      | L CAC   | S CCGS V CCTC C C C K GTC W                          | T<br>3350<br>PCCAG<br>Q<br>3410<br>SCAG<br>S<br>3470<br>AGAAG<br>K<br>3530<br>SGTTT<br>F<br>3590<br>ATCCA   | AAN CGAN E CGG R CCG AT O AT O O O O O O O O O O O O O O O O | CAT<br>I<br>GAA<br>N<br>CCT<br>L<br>CTT<br>F<br>CCA<br>H  | CAF<br>N<br>CAF<br>N<br>CAF<br>K<br>TGT<br>V     | L ATG D ATA I CT C C CT Y           |
| V ACCO P GTTC C ACTT         | 331C TCAACCT N L 337C AGGTCTC V W 3430 CCCTTCCC F P 3490 SCIGCAP C K 3550 TGGGGATC G S 3610 | I  FGCT  K  CCTT  F  AGGI  E  CAGI   | T TGGI V AGTI F TCAI I AGAA K AAGG   | I CCGC A CCCA Q CCGT V AAAA N CAGC       | P CCAT  R CCTT  F CCAT  M        | L COTT F F COGCA A COGGA E TAAA                                   | GGTGTC V C 3330 TGGCTH G Y 3390 CTTCCT F L 3450 TTACTT Y F 3510 GTCTTC S S 3570 TTTCAC F R 3630                   | I ACAC T PGGI V PCTA Y CTGI V                                     | Y CGCA Q CAT M CTG C AAGG             | M<br>GGGG<br>G<br>GGA<br>E<br>CGGT<br>V<br>CTG<br>C | L CACO  | S CCGT V LCTO C C K K GTO W                          | 233500 PCCAG Q Q 34100 PCCAG S 34470 R S 3530 R S 3530 R S S 3530 P S S S S S S S S S S S S S S S S S S S   | EAA  | CAT I GAA N CCT F CCA H GAT I   | CAF<br>N<br>CAF<br>N<br>CAF<br>K<br>TGI<br>V     | L TG D TA I C C C C TG Y SAA S      |
| V ACCO P GTTC C ACTT         | 331C TCAACCT N L 337C AGGTCTCC V W 3430 CCTTCCC F P 3490 SCTGCAP C K 3550 TGGGATC G S 3610  | I  CGCT  K  CCCTT  AGGA  E  CAGA  E  CAGA  E  CAGA  E  CAGA  | T TGGT V AGTT F TCAT I AGGAR K AAGGC | I CCGC A CCCA CCCA V LAAAA N CAGC A      | P CAT R CAT F CAT M CAT I CGGT   | L COTT F F COCC A COCC A TAA N COCC COCC COCC COCC COCC COCC COCC | GGTGTC V C 3330 TGGCTH G Y 3390 CTTCCH F L 3450 TTACTH Y F 3510 GTCTTC S S 3570 TTTCAC F R 3630 GACATO            | I ACAC T PGGI V PCTA Y TGG V CTGG E ECCAC                         | Y CGCA O CAT M CTG C AGG G AGG G RACG | M GGGG G GGA E GGT V CTG C CAT C                    | L CACCATON TO THE CACCATON TO | S CCGC V CCTC C C C C C C C C C C C C C C C C        | T 3350 Q 3410 S S 3470 K S 3530 K T F S 3550 K T C C A F P 3650 C ATGCC | AAN GAAN E CCG R CTG V AT                                    | CAT I GAA N CCT F CCA H GAT I | CAF<br>N CAF<br>N CAF<br>K TGT<br>V TGG<br>G     | L TG D TA I CT C SGT Y SAA S SCT    |
| V ACCO P GTTC C ACTT         | 331C TCAACCT N L 337C AGGTCTCC V W 3430 CCTTCCC F P 3490 SCTGCAP C K 3550 TGGGATC G S 3610  | I  CGCT  K  CCCTT  AGGA  E  CAGA  E  CAGA  E  CAGA  E  CAGA  | T TGGT V AGTT F TCAT I AGGAR K AAGGC | I CCGC A CCCA CCCA V LAAAA N CAGC A      | P CAT R CAT F CAT M CAT I CGGT   | L COTT F F COCC A COCC A TAA N COCC COCC COCC COCC COCC COCC COCC | GGTGTC V C 3330 TGGCTX G Y 3390 CTTCCI F L 3450 TTACTI Y F 3510 GTCTTC S S 3570 TTTCAC F R 3630 GACATC T S        | I  ACAC T  PGGI V  CTGI V  CTGI CTGI CTGI CTGI CTGI CTGI CTGI CTG | Y CGCA O CAT M CTG C AGG G AGG G RACG | M GGGG G GGA E GGT V CTG C CAT C                    | L CACCATON TO THE CACCATON TO | S CCGC V CCTC C C C C C C C C C C C C C C C C        | T 3350 Q 3410 S S 3470 K S 3530 K T F S 3550 K T C C A F P 3650 C ATGCC | AAN GAAN E CCG R CTG V AT                                    | CAT I GAA N CCT F CCA H GAT I | CAF<br>N CAF<br>N CAF<br>K TGT<br>V TGG<br>G     | L TG D TA I CT C SGT Y SAA S SCT    |
| V ACCA  P GTTC C ACTT L GCTC | 331C TCAACCT N L 337C AGGTCTCC V W 3430 CCTTCCC F P 3490 SCTGCAP C K 3550 TGGGATC G S 3610  | I (GCT)  K (CCT)  F (CCT)  CAGG  E (CAGG  CAGG  CAGG  CAGG  G (CAGG  CAGG  CAG | T IGGT V AGTT F ICAT I AGAA K AAGC A | I CCGC A CCGT V AAAA N LAGC A            | P CAT F CAT M CGAT I CGGT V      | L GGTT F GGTA Y CGGC A CGGA E TAA N CTG                           | GGTGTC V C 3330 TGGCTH G Y 3390 CTTCCT F L 3450 TTACTT Y F 3510 GTCTTC S S 3570 TTTCAC F R 3630 GACATC GACATC T S | I  ACAC T  PGGI V  PCTA Y  CTGI V  CTGI T  CCAC                   | Y XSGT V CGCA Q ACAT M CTG C AGG G R  | M GGGG G GGA E GGT V GCTG C GATG C                  | L CAC   | S CCGS V CCTC C C C K W CCTC W T CCAC T T            | CCACC T 3350 PCCAG Q 3410 Q 3410 S 3470 K 3530 K 3530 F F 3590 ATCCA P 3650 CATGC C 3710                    | CATO   | CAT I SAAN N CCT L CTT F CCA H GAT I  | CAF<br>N CAF<br>N CAF<br>K TGI<br>V TGG<br>G CGG | L TG D TA I I CT C C T Y SAA S CT W |

WO 02/010382 PCT/EP01/08309

Fig. 10 / continue on 3

MVGGCRWTEDVEPAEVKEKMSFRAARLSMRNRRNDTLDSTRTLYSSASRSTDLSYSESASFYAAFRTQTCPIMASWDLVNFIQANF
KKRECVFFTKDSKATENVCKCGYAQSQHMEGTQINQSEKWNYKKHTKEFPTDAFGDIQFETLGKKGKYIRLSCDTDAEILYELLTQ
HWHLKTPNLVISVTGGAKNFALKPRMRKIFSRLIYIAQSKGAWILTGGTHYGLMKYIGEVVRDNTISRSSEENIVAIGIAAWGMVS
NRDTLIRNCDAEGYFLAQYLMDDFTRDPLYILDNNHTHLLLVDNGCHGHPTVEAKLRNQLEKYISERTIQDSNYGGKIPIVCFAQG
GGKETLKAINTSIKNKIPCVVVEGSGQIADVIASLVEVEDALTSSAVKEKLVRFLPRTVSRLPEEETESWIKWLKEILECSHLLTV
IKMEEAGDEIVSNAISYALYKAFSTSEQDKDNWNGQLKLLLEWNQLDLANDEIFTNDRRWEKSKPRLRDTIIQVTWLENGRIKVES
KDVTDGKASSHMLVVLKSADLQEVMFTALIKDRPKFVRLFLENGLNLRKFLTHDVLTELFSNHFSTLVYRNLQIAKNSYNDALLTF
VWKLVANFRRGFRKEDRNGRDEMDIELHDVSPITRHPLQALFIWAILQNKKELSKVIWEQTRGCTLAALGASKLLKTLAKVKNDIN
AAGESEELANEYETRAVGESTVWNAVVGADLPCGTDIASGTHRPDGGELFTECYSSDEDLAEQLLVYSCEAWGGSNCLELAVEATD
QHFIAQPGVQNFLSKQWYGEISRDTKNWKIILCLFIIPLVGCGFVSFRKKPVDKHKKLLWYYVAFFTSPFVVFSWNVVFYIAFLLL
FAYVLLMDFHSVPHPPELVLYSLVFVLFCDEVRQGRPAAPSAGPAKPTPTRNSIWPASSTRSPGSRSRHSFHTSLQAEGASSGLGQ
PRKGWTFKNLEMVDISKLLMSLSVPFCTQWYVNGVNYFTDLWNVMDTLGLFYFIAGIVFRQGILRQNEQRWRWIFRSVIYEPYLAM
FGQVPSDVDGTTYDFAHCTFTGNESKPLCVELDEHNLPRFPEWITIPLVCIYMLSTNILLVNLLVAMFGYTVGTVQENNDQVWKFQ
RYFLVQEYCSRLNIPFFFIVFAYFYMVVKKCFKCCCKEKNMESSVCCEWFIHVYLGSEAAINFREGCLHPVIGSWTPGWLVWTSTR
ILTCSAGWPAAGSLSVTTHSSWVPAKSSKSQAHPDRTGRECDSASGWEGQPARWVEESVALFGHRGPVWPPTTLGITELNAPVL

В.

O L 2310 TGCTGGTCTATTCCTGTGAAGCTTGGGGTGGAAGCAACTGTCTGGAGCTGGCGGTGGAGG LVYSCEAWGGSNCLELAVEA 2350 2370 TDQHFIAQPGVQNFLSKQWY 2430 2450 ATGGAGAGATTTCCCGAGACACCAAGAACTGGAAGATTATCCTGTGTCTGTTTATTATAC G E I S R D T K N W K I I L C L F I I P 2490 2510 CCTTGGTGGGCTGTGGCTTTGTATCATTTAGGAAGAACCTGTCGACAAGCACAAGAAGC LVGCGFVSFRKKPVDK

Figure 11:

a.) Trp10b cDNA and derived amino acid sequence

|  |   | 10   |  |  |  |   |  | 30   |  |  |   |  |   | 50  |                                       |                                       |  |
|--|---|--|--|--|--|---|--|--|--|--|---|--|---|---|---------------------------------------|---------------------------------------|--|
| ΑŢ   | GAA   | ATCCI  | TCC  | rtcc   | TGT  | 'CCA  | CAC  | CATCG  | rgc'i  | TAT                                    | CAG   | GGA  | GA  | TGT   | GTG                                   | CAA                                   | GTGT   |
| M  | K   | S F  | _  | P  | v  | H   | T  | I V<br>90  | L  | I                                      | R   | E  | N   | V<br>110  | C                                     | K                                     | С  |
| ac.  | עיזוי <i>ף</i>                                    |  |  | 7007   | 001  | ~~~   | 277713                                       |  |  |  |   | ~~*  | n n c   |   | ~                                     | 3 mc                                  |  |
|  |   |  |  |  |  |   |  | AGGCA  |  |  |   |  |   |   |                                       |                                       |  |
| G  | Y   | A Q  |  | Q  | H  | M   | E  |  | Q  | 1                                      | N   | Q  | S   |   | K                                     | W                                     | Ŋ  |
|  |   | 130  |  |  |  |   |  | 150  |  |  | •   |  |   | 170   |                                       |                                       |  |
| TA   | CAA   |  |  | CCAA   | LGGA   | ATT   | TCC  | TACCG  |  |  | TGG   | GGA  | TAT   | CTCA  | GTT                                   | TGA                                   | GACA   |
| Y  | K   | K H  | T  | K  | E  | F   | P  | T D  | Α  | F                                      | G   | Д  | I   | Q   | F                                     | E                                     | ${f T}$  |
|  |   | 190  | 1  |  |  |   |  | 210  |  |  |   |  |   | 230   |                                       |                                       |  |
| CI   | 'GGG  | GAAGA  | AAGG   | <b>GAA</b>   | GTA  | TAT   | ACG  | TCTGT  | CTC  | CGA                                    | CAC   | GGA  | .CGC  | CGGA  | TAA                                   | CCT                                   | TTAC   |
| L  | G   | K K  | G  | K  | Ÿ  | I   | R  | L S  | С  | D                                      | T   | D  | А   | E   | 1                                     | L                                     | Y  |
|  |   | 250  | _  |  |  | _   |  | 270  | -  | _                                      | _   | _  |   | 290   | _                                     |                                       | -  |
| GA   | GCT   |  |  | מכים   | CTY  | GCA   | י כי בי                                      | GAAAAC   | יאמיר  | ממיזי                                  | רכידי   | יכפייי                                     | ראיז  |   | ጥርታጥ                                  | የመር                                   | rccc   |
| E  | Ĺ   | LI   |  | H  | W  | H   | L  |  |  |  |   | V.   |   | s   |                                       | T                                     | G  |
|  |   | 310  | ~  | 41   | **   | -11   | 13   | 330  | F  | TA                                     | n   | ٧.   | _   | 350   | v                                     | 1                                     | G  |
| ~~   | -   |  |  |  |  | ~~~   | ~~~  |  | .~~~   | ~                                      | ~>  | ~===                                       |   |   | ~ ~~                                  |                                       |  |
| _  |   |  |  |  |  |   |  | GCGCAT   |  |  |   |  |   |   |                                       |                                       |  |
| G  | A   | K N  | _  | A  | L  | K   | P  |  | R  | K                                      | 1   | F  | S   |   | L                                     | .I                                    | Y  |
|  |   | 370  |  |  |  |   |  | 390  |  |  |   |  |   | 410   |                                       |                                       |  |
| ΤA   | CGC   | GCAGI  | 'CCA   | AGG  | TGC  | TTG   | GAT  | TCTCAC   | 'GGG   | AGG                                    | CAC   | CCA  | TTZ   | ATGG  | CCI                                   | GAT                                   | GAAG   |
| I  | A   | Q S  | K  | G  | Α  | W   | I  | L T  | G  | G                                      | T   | H  | Y   | G   | L                                     | М                                     | K  |
|  |   | 430  |  |  |  |   |  | 450  |  |  |   |  |   | 470   |                                       |                                       |  |
| TA   | CAT   | CGGGG  | AGG1   | GGT  | GAG  | AGA   | AAT  | CACCAI   | 'CAG   | CAG                                    | GAG   | TTC  | AGZ   | AGGA  | GAA                                   | TAT                                   | TGTG   |
| Y  | I   | G E  | v  | V  | R  | D   | N  | TI   | S  | R                                      | s   | s  | E   | E   | N                                     | I                                     | V  |
|  |   | 490  |  |  |  |   |  | 510  |  |  |   |  |   | 530   |                                       |                                       |  |
| GC   | CAT   | TGGCA  | TAGO   | AGC  | TTG  | GGG   | CAT  | GGTCTC   | מבים:  | CCG                                    | GGA   | CAC  | CCI   | יייבאייי  | CAG                                   | GAA                                   | ሞግርር   |
| A  |   | G I  |  | A  |  |   | М  |  |  |  | D   |  |   | I   | -                                     | N                                     | C  |
|  | _   | 550  |  |  | ••   | _   | ••   | 570  |  |  | 2   |  |   | 590   |                                       | 1.                                    | Č  |
| C D  | ייטיטיי   |  |  | . HATEL  | alatara.   | 700   | יררא   | GTACCI   | ייי א פייי   | ערטא                                   | ת רינוני  |  | ~~~   |   | ארי א                                 | maa                                   | א כייינים  |
| D  |   | E G  |  | F  | 111.<br>L  |   |  |  |  |  |   |  |   |   |                                       |                                       | AC1G   |
| ע  | A   | B 6  | Y  | r  | ш  | A   | Q  | Y L  | M  | D                                      | D   | F  | T   | R   | D                                     | P                                     | ш  |
|  |   |  |  |  |  |   |  |  |  |  |   |  |   |   |                                       |                                       |  |
|  |   | 610  |  |  |  |   |  | 630  |  |  |   |  |   | 650   |                                       |                                       |  |
|  |   | CCTGG  | ACAA   |  |  |   |  | TTTGCT   |  |  |   |  |   | CTG'  |                                       | TGG                                   |  |
| TA<br>Y  | TAT(  |  | ACAA   | CAA<br>N   | CCA<br>H   | CAC<br>T  | ACA<br>H                                     |  | GCT<br>L   |  |   | CAA<br>N                                   |   |   | TCA<br>H                              | TGG<br>G                              | ACAT<br>H  |
|  |   | CCTGG  | ACAA   |  |  |   |  | TTTGCT   |  |  |   |  |   | CTG'  |                                       |                                       |  |
| Y  | I   | CCTGG<br>L D<br>670  | ACAA<br>N  | N  | Ħ  | T   | H  | TTTGCT<br>L L  | L  | V                                      | D   | N  | G   | CTG'<br>C<br>710  | H                                     | G                                     | H  |
| Y  | I<br>CAC'   | CCTGG<br>L D<br>670  | ACAA<br>N<br>AAGC                                | N  | Ħ  | T   | H  | TTTGCT<br>L L<br>690<br>TCAGCT   | L  | V                                      | D   | N  | G   | CTG<br>C<br>710<br>CTGA   | H                                     | G                                     | H<br>TATT  |
| Y<br>CC  | I<br>CAC'   | CCTGG<br>L D<br>670<br>TGTCG   | ACAA<br>N<br>AAGC                                | n<br>'AAA  | H<br>GCT   | T   | H<br>GAA                                     | TTTGCT<br>L L<br>690<br>TCAGCT   | L<br>AGA   | V<br>Gaa                               | D<br>GTA  | N<br>TAT                                   | G<br>CTC  | CTG<br>C<br>710<br>CTGA   | H<br>GCG                              | G<br>CAC                              | H<br>TATT  |
| Y<br>CC<br>P   | I<br>CAC'<br>T                                    | CCTGG<br>L D<br>670<br>TGTCG<br>V E<br>730   | ACAA<br>N<br>AAGO<br>A                           | N<br>'AAA<br>K                                       | H<br>GCT<br>L                                    | T<br>CCG<br>R                                     | H<br>GAA<br>N                                | TTTGCT<br>L L<br>690<br>TCAGCT<br>Q L<br>750   | L<br>AGA<br>E                                      | V<br>GAA<br>K                          | D<br>GTA<br>Y   | N<br>TAT<br>I                              | G<br>CTC<br>S   | CTG'<br>C<br>710<br>TGAC<br>E<br>770  | H<br>GCG<br>R                         | G<br>CAC<br>T                         | H<br>TATT<br>· I                                 |
| Y<br>CC<br>P<br>CA   | I<br>CAC'<br>T<br>AGA'                            | CCTGG L D 670 TGTCG V E 730 TTCCA  | ACAA<br>N<br>AAGC<br>A                           | N<br>'AAA<br>K<br>.TGG                               | H<br>GCT<br>L<br>TGG                             | T<br>CCG<br>R<br>CAA                              | H<br>GAA<br>N<br>GAT                         | TTTGCT L L 690 TCAGCT Q L 750 CCCCAT   | L<br>AGA<br>E<br>TGT                               | V<br>GAA<br>K<br>GTG                   | D<br>GTA<br>Y<br>TTT  | n<br>Tat<br>I<br>TGC                       | G<br>CTC<br>S<br>CCA  | CTG'<br>710<br>TGAG<br>E<br>770   | H<br>GCG<br>R<br>AGG                  | G<br>CAC<br>T<br>TGG                  | H<br>TATT<br>· I<br>AAAA                         |
| Y<br>CC<br>P   | I<br>CAC'<br>T                                    | CCTGG L D 670 TGTCG V E 730 TTCCA  | ACAA<br>N<br>AAGC<br>A                           | N<br>'AAA<br>K                                       | H<br>GCT<br>L                                    | T<br>CCG<br>R<br>CAA                              | H<br>GAA<br>N                                | TTTGCT L L 690 TCAGCT Q L 750 CCCCAT P I   | L<br>AGA<br>E<br>TGT                               | V<br>GAA<br>K                          | D<br>GTA<br>Y<br>TTT  | n<br>Tat<br>I<br>TGC                       | G<br>CTC<br>S<br>CCA  | CTGAC<br>TGAC<br>E<br>770<br>AAGGA  | H<br>GCG<br>R<br>AGG                  | G<br>CAC<br>T                         | H<br>TATT<br>· I                                 |
| Y<br>CC<br>P<br>CA   | I<br>CAC'<br>T<br>AGA'<br>D                       | CCTGG L D 670 TGTCG V E 730 TTCCA S N 790  | ACAA<br>N<br>AAGC<br>A<br>ACTA<br>Y              | n<br>'AAA<br>K<br>.TGG<br>G                          | H<br>GCT<br>L<br>TGG<br>G                        | T<br>CCG<br>R<br>CAA<br>K                         | H<br>GAA<br>N<br>GAT                         | TTTGCT L L 690 TCAGCT Q L 750 CCCCAT P I 810   | L<br>AGA<br>E<br>TGT<br>V                          | V<br>GAA<br>K<br>GTG<br>C              | D<br>GTA<br>Y<br>TTT<br>F                                   | N<br>TAT<br>I<br>TGC<br>A                  | G<br>CTC<br>S<br>CCA<br>Q   | CTG' 710 CTGAG E 770 AGGG   | H<br>GCG<br>R<br>AGG<br>G             | G<br>CAC<br>T<br>TGG<br>G             | H<br>TATT<br>· I<br>AAAA<br>K                    |
| Y<br>CC<br>P<br>CA<br>Q<br>GA                                    | I<br>CAC'<br>T<br>AGA'<br>D<br>GAC'               | CCTGG L D 670 TGTCG V E 730 TTCCA S N 790 TTTGA  | ACAA<br>N<br>AAGC<br>A<br>ACTA<br>Y              | N<br>K<br>TGG<br>G                                   | H<br>GCT<br>L<br>TGG<br>G<br>G                   | T<br>CCG<br>R<br>CAA<br>K<br>TAC                  | H<br>GAA<br>N<br>GAT<br>I<br>CTC             | TTTGCT L L 690 TCAGCT Q L 750 CCCCAT P I 810 CATCAA  | L<br>PAGA<br>E<br>TGT<br>V                         | V<br>GAA<br>K<br>GTG<br>C              | D<br>GTA<br>Y<br>TTT<br>F<br>AAT                            | N<br>TAT<br>I<br>TGC<br>A                  | G<br>CTC<br>S<br>CCA<br>Q   | CTG' 710 TTGAG E 770 AAGGA G 830  | H<br>GCG<br>R<br>AGG<br>G<br>GGT      | G<br>CAC<br>T<br>TGG<br>G<br>G        | H TATT I AAAA K GGAA                             |
| Y<br>CC<br>P<br>CA   | I<br>CAC'<br>T<br>AGA'<br>D<br>GAC'               | CCTGG L D 670 TGTCG V E 730 TTCCA S N 790 ITTGA L K  | ACAA<br>N<br>AAGC<br>A<br>ACTA<br>Y              | N<br>K<br>TGG<br>G                                   | H<br>GCT<br>L<br>TGG<br>G                        | T<br>CCG<br>R<br>CAA<br>K<br>TAC                  | H<br>GAA<br>N<br>GAT<br>I<br>CTC             | TTTGCT L L 690 TCAGCT Q L 750 CCCCAT P I 810 CATCAA I K  | L<br>PAGA<br>E<br>TGT<br>V                         | V<br>GAA<br>K<br>GTG<br>C              | D<br>GTA<br>Y<br>TTT<br>F<br>AAT                            | N<br>TAT<br>I<br>TGC<br>A                  | G<br>CTC<br>S<br>CCA<br>Q   | CTG' 710 CTGAC E 770 AGGZ G 830 CTGTC   | H<br>GCG<br>R<br>AGG<br>G             | G<br>CAC<br>T<br>TGG<br>G             | H<br>TATT<br>· I<br>AAAA<br>K                    |
| Y<br>CC<br>P<br>CA<br>Q<br>GA<br>E                               | I<br>CAC'<br>T<br>AGA'<br>D<br>GAC'               | CCTGG L D 670 TGTCG V E 730 TTCCA S N 790 TTTGA L K 850  | ACAA<br>N<br>AAGC<br>A<br>ACTA<br>Y<br>AAGC<br>A | N<br>K<br>TGG<br>G<br>CAT                            | H<br>GCT<br>L<br>TGG<br>G<br>G<br>CAA'           | T<br>CCG<br>R<br>CAA<br>K<br>TAC                  | H<br>GAA<br>N<br>GAT<br>I<br>CTC             | TTTGCT L L 690 TCAGCT Q L 750 CCCCAT P I 810 CATCAA I K 870  | L<br>AGA<br>E<br>TGT<br>V<br>AAA<br>N              | V<br>GAA<br>K<br>GTG<br>C<br>TAA       | D<br>GTA<br>Y<br>TTT<br>F<br>AAT                            | N<br>TAT<br>I<br>TGC<br>A<br>TCC           | G<br>CTC<br>S<br>CC#<br>Q<br>TTC  | CTG' 710 CTGAC E 770 AAGGA G 830 CTGTC V 890  | H<br>GCG<br>R<br>AGG<br>G<br>GGT<br>V | G<br>T<br>TGG<br>G<br>V               | H TATT · I AAAA K GGAA E                         |
| Y CC P CA Q GA GA GG GG GG                                       | I<br>CAC'<br>T<br>AGA'<br>D<br>GAC'<br>T          | CCTGG L D 670 TGTCG V E 730 TTCCA S N 790 TTTGA L K 850 3GGCC  | ACAA<br>N<br>AAGC<br>ACTA<br>Y<br>AAGC<br>A      | N AAA K TGG G CAT                                    | H<br>GCT<br>L<br>TGG<br>G<br>CAA'<br>N           | T<br>CCG<br>R<br>CAA<br>K<br>TAC<br>T             | H<br>GAA<br>N<br>GAT<br>I<br>CTC<br>S        | TTTGCT L L 690 TCAGCT Q L 750 CCCCAT P I 810 CATCAA I K 870 CGCTAG   | AGA TGT V AAA N CCT                                | GAA<br>K<br>GTG<br>C<br>TAA<br>K       | D<br>GTA<br>Y<br>TTT<br>F<br>AAT<br>I                       | TAT I TGC A TCC P                          | G<br>CTC<br>S<br>CCA<br>Q<br>TTC<br>C   | CTG' 710 CTGAG E 770 AAGGA G 830 CTGTG V 890  | H<br>GCG<br>R<br>AGG<br>G<br>GGT<br>V | G<br>T<br>TGG<br>G<br>GGT<br>V        | H TATT I AAAA K GGAA E GACA                      |
| Y CC P CA Q GA GA GG GG GG                                       | I<br>CAC'<br>T<br>AGA'<br>D<br>GAC'<br>T          | CCTGG L D 670 TGTCG V E 730 TTCCA S N 790 TTTGA L K 850 3GGCC  | ACAA<br>N<br>AAGC<br>ACTA<br>Y<br>AAGC<br>A      | N AAA K TGG G CAT                                    | H<br>GCT<br>L<br>TGG<br>G<br>CAA'<br>N           | T<br>CCG<br>R<br>CAA<br>K<br>TAC<br>T             | H<br>GAA<br>N<br>GAT<br>I<br>CTC<br>S        | TTTGCT L L 690 TCAGCT Q L 750 CCCCAT P I 810 CATCAA I K 870 CGCTAG A S   | AGA TGT V AAA N CCT                                | GAA<br>K<br>GTG<br>C<br>TAA<br>K       | D<br>GTA<br>Y<br>TTT<br>F<br>AAT<br>I                       | TAT I TGC A TCC P                          | G<br>CTC<br>S<br>CCA<br>Q<br>TTC<br>C   | CTG' 710 CTGAG E 770 AAGGA G 830 CTGTG V 890  | H<br>GCG<br>R<br>AGG<br>G<br>GGT<br>V | G<br>T<br>TGG<br>G<br>GGT<br>V        | H TATT I AAAA K GGAA E GACA                      |
| Y CC P CA Q GA GA GG GG GG                                       | I<br>CAC'<br>T<br>AGA'<br>D<br>GAC'<br>T          | CCTGG L D 670 TGTCG V E 730 TTCCA S N 790 TTTGA L K 850 3GGCC  | ACAA<br>N<br>AAGC<br>ACTA<br>Y<br>AAGC<br>A      | N AAA K TGG G CAT                                    | H<br>GCT<br>L<br>TGG<br>G<br>CAA'<br>N           | T<br>CCG<br>R<br>CAA<br>K<br>TAC<br>T             | H<br>GAA<br>N<br>GAT<br>I<br>CTC<br>S        | TTTGCT L L 690 TCAGCT Q L 750 CCCCAT P I 810 CATCAA I K 870 CGCTAG   | AGA TGT V AAA N CCT                                | GAA<br>K<br>GTG<br>C<br>TAA<br>K       | D<br>GTA<br>Y<br>TTT<br>F<br>AAT<br>I                       | TAT I TGC A TCC P                          | G<br>CTC<br>S<br>CCA<br>Q<br>TTC<br>C   | CTG' 710 CTGAG E 770 AAGGA G 830 CTGTG V 890  | H<br>GCG<br>R<br>AGG<br>G<br>GGT<br>V | G<br>T<br>TGG<br>G<br>GGT<br>V        | H TATT I AAAA K GGAA E GACA                      |
| Y CC P CA Q GA G G G G G G G G G G G G G G G G G                 | I<br>CAC'<br>T<br>AGA'<br>D<br>GAC'<br>T<br>CTCC  | CCTGG L D 670 TGTCG V E 730 TTCCA S N 790 TTTGA L K 850 GGCC G Q   | ACAA<br>N<br>AAGC<br>ACTA<br>Y<br>AAGC<br>A      | N<br>K<br>TGG<br>G<br>CAT<br>I                       | H<br>GCTG<br>L<br>TGGG<br>G<br>CAA'<br>N<br>TGA' | T<br>CCG<br>R<br>CAA<br>K<br>TAC<br>T<br>T<br>TGT | H<br>GAA<br>N<br>GAT<br>I<br>CTC<br>S<br>GAT | TTTGCT L L 690 TCAGCT Q L 750 CCCCAT P I 810 CATCAA I K 870 CGCTAG A S 930   | L<br>PAGA<br>E<br>TGT<br>V<br>AAA<br>N<br>CCT<br>L | V GAA K GTG C TAA K GGT                | D<br>GTA<br>Y<br>TTT<br>F<br>AAT<br>I<br>GGA                | TATI TGC A TCC P GGT V                     | G<br>CTC<br>S<br>CCA<br>Q<br>TTC<br>C   | CTG' C 710 CTGAC E 770 AGGAC G 830 CTGTC V 890 AGGAC D 950                                | H<br>GCG<br>R<br>AGG<br>G<br>GGT<br>V | G<br>T<br>TGG<br>G<br>GGT<br>V<br>CCT | H TATT I AAAA K GGAA E GACA T                    |
| Y CC P CA Q GA E GG G TC   | I CACT AGAT D GACT T CTCC                         | CCTGG L D 670 TGTCG V E 730 TTCCA S N 790 TTTGA L K 850 GGCC G Q 910   | ACAA N AAGC A AAGC A AGAT I                      | N AAA K TGG G CAT I CGC A                            | H GCT L TGG G CAA' N TGA' D                      | T CCGG R CAA K TAC T T GT V                       | H GAA N GAT I CTC S GAT GGT GGT              | TTTGCT L L 690 TCAGCT Q L 750 CCCAT P I 810 CATCAA I K 870 CGCTAG A S 930 GCGCTT   | L AGA E TGT V AAAA N CCT L                         | V GAA K GTG C TAA K GGT V              | GTA' Y TTT F AAT I GGAG                                     | N TAT I TGC A TCC P GGT V                  | G CTC S CCA Q TTC C C GGA E GGT   | CTG' C 710 CTGAC E 770 AGGA G 830 CTGTC V 890 CGGA D 950                                  | H GCG R AGG G V TGC A                 | G CAC T TGG G G CCT L GCT             | H TATT I AAAA K GGAA E GACA T GCCT               |
| Y CC P CA Q GA E GG G TC   | I CACT AGAT D GACT T CTCC                         | CCTGG L D 670 TGTCG V E 730 TTCCA S N 790 TTTGA L K 850 GGCC G Q 910 TGCCG A V                                 | ACAA N AAGC A AAGC A AGAT I                      | N AAA K TGG G CAT I CGC A                            | H GCT L TGG G CAA' N TGA' D                      | T CCGG R CAA K TAC T T GT V                       | H GAA N GAT I CTC S GAT GGT GGT              | TTTGCT L L 690 TCAGCT Q L 750 CCCAT P I 810 CATCAA I K 870 CGCTAG A S 930 GCGCTT R F                                       | L AGA E TGT V AAAA N CCT L                         | V GAA K GTG C TAA K GGT V              | GTA' Y TTT F AAT I GGAG                                     | N TAT I TGC A TCC P GGT V                  | G CTC S CCA Q CTTC C C GGA E GGT V  | CTGACTGACTGACTGACTGACTGACTGACTGACTGACTGA  | H GCG R AGG G V TGC A                 | G CAC T TGG G G CCT L GCT             | H TATT I AAAA K GGAA E GACA T GCCT               |
| Y CC P CA Q GA E GG G TC S                                       | I CAC' T AGA' D GAC' T CTCC S                     | CCTGG L D 670 TGTCG V E 730 TTCCA S N 790 TTTGA L K 850 GGCC G Q 910 TGCCG A V                                 | ACAA N AAGC A AAGC A AGAT I ICAA K               | N CAAA K TGG G CAT I CGC A GGAA                      | H GCT' L TGGG G CAA' N TGA' D GAAA               | T CCG R CAAA K TACC T T GTC V GCTC L              | H GAA N GATT I CTC S GATT I GGTT V           | TTTGCT L L 690 TCAGCT Q L 750 CCCAT P I 810 CATCAA I K 870 CGCTAG A S 930 GCGCTT R F                                       | L AGA E TGT V AAAA N CCT L TTT                     | V GAAA K GTG C TAAA K GGT V ACC P      | D GTA Y TTT F AAT I GGA( E CCGG R                           | N TAT I TGC A TCC P GGT V CACC T           | G CTC S CCA Q CTC C C GGA E GGT V 1   | C 710<br>710<br>710<br>E 7770<br>AAGGA<br>830<br>V 890<br>AGGA<br>D 950<br>CGTCC<br>S 010 | H GCG R AGG G GGT V TGC A CCG R       | G CAC T TGG G G CT L                  | H TATT I AAAA K GGAA E GACA T GCCT P             |
| Y CC P CA Q GA E GG G TC S GA G GA G G G G G G G G G G G G G G G | I CAC' T AGA: D GAC' T CTCC S TTCT S              | CCTGG L D 670 TGTCG V E 730 TTCCA S N 790 TTTGA L K 850 GGCC G Q 910 TGCCG A V 970 GGAGAG                      | ACAA N AAGC A AGAT AGAT I CCAA K CTGA            | N CAAA K TGG G CAT I CGC A GGAA E GAG                | H GCT L TGGG G CAA' N TGAA D GAAG K              | T CCG R CAAA K TACC T GT V GCT L GATC             | H GAA N GAT I CTCC S GAT I GGT V CAA         | TTTGCT L L 690 TCAGCT Q L 750 CCCAT P I 810 CATCAA I K 870 CGCTAG A S 930 GCGCTT R F 990 ATGGCT                            | L AGA E TGT V AAAA N CCT L TTT L                   | V GAAA K GTG C TAAA K GGT V ACC P      | D GTA Y TTTT F AATT I GGA E CCGG R                          | N TAT I TGC A TCC P CACC T                 | G CTC S CCA Q C C C C C C C C C C C C C C C C C   | C 710 710 7710 7710 7710 7710 7710 7710 7   | H SCG R AGG G SGT V TGC A CCG R       | G CAC T TGG G GGT V CCT L GCT L TCA   | H TATT I AAAA K GGAA E GACA T GCCT P             |
| Y CC P CA Q GA E GG G TC S                                       | I CAC' T AGA: D GAC' T CTCC S TTCT S              | CCTGG L D 670 TGTCG V E 730 TTCCA S N 790 TTTGA L K 850 GGCC G Q 910 TGCCG A V 970 GGAGAG E T                  | ACAA N AAGC A AGAT AGAT I CCAA K CTGA            | N CAAA K TGG G CAT I CGC A GGAA E GAG                | H GCT L TGGG G CAA' N TGA: D GAAG K              | T CCG R CAAA K TACC T GT V GCT L GATC             | H GAA N GAT I CTC S GAT V CAAA               | TTTGCT L L 690 TCAGCT Q L 750 CCCAT P I 810 CATCAA I K 870 CGCTAG A S 930 GCGCTT R F 990 ATGGCT W L                        | L AGA E TGT V AAAA N CCT L TTT L                   | V GAAA K GTG C TAAA K GGT V ACC P      | D GTA Y TTTT F AATT I GGA E CCGG R                          | N TAT I TGC A TCC P CACC T                 | G CTC S CCA Q CTC C C GGA E GGT V 1 CGGA E CGGA E   | C C C C C C C C C C C C C C C C C C C   | H SCG R AGG G SGT V TGC A CCG R       | G CAC T TGG G GGT V CCT L GCT L TCA   | H TATT I AAAA K GGAA E GACA T GCCT P             |
| Y CCA P CAA Q GAG G TCC S GAG E                                  | I CAC' T AGA: D GAC' T CTCC S FTTCT S E GGGGG     | CCTGG L D 670 TGTCG V 730 TTCCA S N 790 TTTGA L K 850 GGCC G Q 910 TGCCG A V 970 GGAGAG E T                    | ACAA N AAGC A AAGC A AGAT I CAA K CTGA           | N  CAAA  K  CTGG  CCAT  I  CCGC  A  GGAA  E  GAGS    | H GCT' L TTGGG G CAA' N TTGA' D SAAC K W         | T CCGG R CAAC K TACC T U GCTC L GATC I            | H GAA N GAT I CTC S GAT I GGT V CAAA         | TTTGCT L L 690 TCAGCT Q L 750 CCCCAT P I 810 CATCAA I K 870 CGCTAG A S 930 GCGCTT R F 990 ATGGCT W L                       | L AGA E TGT V AAA N CCT L TTT L CAA                | V GAA K GTG C TAA K GGT V ACC P AGA E  | D GTA Y TTTT F AAT I GGA E CCGG R AAT I                     | N TAT I TGC A TCC P CAC T TCT L            | G CTC S CCA Q Q TTC C G G A E G G G A E G G G A E G G | C C C C C C C C C C C C C C C C C C C   | H SCG R AGG G G TGC A CCG R TTC S     | G CAC T TGG G GGT V CCT L GCT L TCA   | H TATT I AAAA K GGAA E GACA T GCCT P CCTA L      |
| Y CCC P CAL Q GAG E GGG G TCC S GAG E                            | I CAC' T AGA: D GAC' T CTCC S FITCI S AGAC E AACA | CCTGG L D 670 TGTCG V E 730 TTCCA S N 790 TTTGA L K 850 GGCC G Q 910 TGCCG A V 970 GGAGAG E T 1030             | ACAA N AAGC A AGAT I CAA K CTGA E                | N  CAAA  K  CTGG  CCAT  I  CCGC  A  GGAA  E  GAAC  S | H GCT' L TGGG G CAA' N TGA' D GAAG K W GGGA      | T CCG R CAA K TAC T T GT V GGT L GAA I            | H GAA N GAT I CTC S GAT I CAA K K AGC        | TTTGCT L L 690 TCAGCT Q L 750 CCCCAT P I 810 CATCAA I K 870 CGCTAG A S 930 GCGCTT R F 990 ATGGCT W L 1050 IGGGGA           | L AGA E TGT V AAAA N CCT L TTTT L CAA K            | V GAA K GTG C TAA K GGT V ACC P AGA E  | D GTA Y TTTT F AAT I GGA E CCGG R AAT I                     | N TAT I TGC A TCC P CACC T CACC T L GAGG   | G CTC S CCA Q Q TTC C GGA E GGA E 1 CGA A   | C C C C C C C C C C C C C C C C C C C   | H GCG G G GCCG R TTC S CAT            | G CAC T TGG G G G CCT L G CTCA H CTCC | H TATT I AAAA K GGAA E GACA T GCCT P CCTA L      |
| Y CCC P CAL Q GAG E GGG G TCC S GAG E                            | I CAC' T AGA: D GAC' T CTCC S FITCI S AGAC E AACA | CCTGG L D 670 TGTCG V E 730 TTCCA S N 790 TTTGA L K 850 GGCC G Q 910 TGCCG A V 970 GGAGAG E T 1030             | ACAA N AAGC A AGAT I CAA K CTGA E                | N  CAAA  K  CTGG  CCAT  I  CCGC  A  GGAA  E  GAAC  S | H GCT' L TGGG G CAA' N TGA' D GAAG K W GGGA      | T CCG R CAA K TAC T T GT V GGT L GAA I            | H GAA N GAT I CTC S GAT I CCAA K AGC A       | TTTGCT L L 690 TCAGCT Q L 750 CCCCAT P I 810 CATCAA I K 870 CGCTAG A S 930 GCGCTT R F 990 ATGGCT W L 1050 IGGGGA G D       | L AGA E TGT V AAAA N CCT L TTTT L CAA K            | V GAA K GTG C TAA K GGT V ACC P AGA E  | D GTA Y TTTT F AAT I GGA E CCGG R AAT I                     | N TAT I TGC A TCC P CACC T T CACC T L GAGG | G CTC S CCA Q C C GGA E C GGA N C C C C C C C C C C C C C C C C C C   | C C C C C C C C C C C C C C C C C C C   | H GCG G G GCCG R TTC S CAT            | G CAC T TGG G G G CCT L G CTCA H CTCC | H TATT I AAAA K GGAA E GACA T GCCT P CCTA L      |
| Y CCC P CAL Q GAG E GGG G TCC S GAG E TTI                        | I CAC' T AGA: D GAC' T CTCC S FTCT S AACA T       | CCTGG L D 670 TGTCG V 730 TTCCA S N 790 TTTGA L K 850 GGCC G Q 910 TGCCG A V 970 GGAGA E T 1030 AGTTA V I 1090 | ACAA N AAGC A AGAT I CAA K CTGA E TTAA K         | N TAAA K TTGG G CAT I CCGC A GGAA E GAAC S           | H GCT' L TGGG G CAA' N TGA' D GAAG K W GGAA E    | T CCG R CAA K TAC T T GT V GCT L GAT E            | H GAA N GAT I CTC S GAT I GGT V CCAA K AGC A | TTTGCT L L 690 TCAGCT Q L 750 CCCCAT P I 810 CATCAA I K 870 CGCTAG A S 930 GCGCTT R F 990 ATGGCT W L 1050 IGGGGA' G D 1110 | L AAGA E TGT V AAAA N CCT L TTT L CAAA K E         | V GAA K GTG C TAA K V ACC P AGA E AATT | D GTA Y TTT F AAT I GGA C C C C C C C C C C C C C C C C C C | N TAT I TGC A TCC P CACC T TCTC L GAGG     | G CTC S CCA Q C C GGA E C GGA N 1   | C C C C C C C C C C C C C C C C C C C   | H GCG R GGT V TGC A CCG R TTC         | G CAC T TGG G GGT V CCT L TCA H CTCA  | H TATT I AAAA K GGAA E GACA T GCCT P CCTA L CTAC |
| Y CCC P CAL Q GAG E GGG G TCC S GAG E TTI                        | I CAC' T AGA: D GAC' T CTCC S FTCT S AACA T       | CCTGG L D 670 TGTCG V 730 TTCCA S N 790 TTTGA L K 850 GGCC G Q 910 TGCCG A V 970 GGAGA E T 1030 AGTTA V I 1090 | ACAA N AAGC A AGAT I CAA K CTGA E TTAA K         | N TAAA K TTGG G CAT I CCGC A GGAA E GAAC S           | H GCT' L TGGG G CAA' N TGA' D GAAG K W GGAA E    | T CCG R CAA K TAC T T GT V GCT L GAT E            | H GAA N GAT I CTC S GAT I GGT V CCAA K AGC A | TTTGCT L L 690 TCAGCT Q L 750 CCCCAT P I 810 CATCAA I K 870 CGCTAG A S 930 GCGCTT R F 990 ATGGCT W L 1050 IGGGGA G D       | L AAGA E TGT V AAAA N CCT L TTT L CAAA K E         | V GAA K GTG C TAA K V ACC P AGA E AATT | D GTA Y TTT F AAT I GGA C C C C C C C C C C C C C C C C C C | N TAT I TGC A TCC P CACC T TCTC L GAGG     | G CTC S CCA Q C C GGA E C GGA N 1   | C C C C C C C C C C C C C C C C C C C   | H GCG R GGT V TGC A CCG R TTC         | G CAC T TGG G GGT V CCT L TCA H CTCA  | H TATT I AAAA K GGAA E GACA T GCCT P CCTA L CTAC |

Fig. 11 (Continuation)

|  |  | 2410 2430 |        |              |        |         |     |      |              |      |         |        | 2450     |       |      |          |       |         |            |  |  |
|--|--|-----------|--------|--------------|--------|---------|-----|------|--------------|------|---------|--------|----------|-------|------|----------|-------|---------|------------|--|--|
| AGAAACTTAGGACCCAAGATTATAATGCTGCAGAGGATGCTGATCGATGTGTTCTTCTTR N L G P K I I M L Q R M L I D V F F F |  |           |        |              |        |         |     |      |              |      |         |        |          | CTTC  |      |          |       |         |            |  |  |
| R  | N  | L         | G      | P            | K      | I       | I   | M    | L            | Q    | R       | M      | L        | I     | D    | V        | F     | F       | F          |  |  |
| 2470 2490 2510 CTGTTCCTCTTTGCGGTGTGGATGGTGGCCTTTGGCGTGGCCAGGCAAGGGATCCTT                           |  |           |        |              |        |         |     |      |              |      |         |        |          |       |      |          |       |         |            |  |  |
| CT   | GTT  | CCT       | CTT    | rgc          | GGT    | GTG     | GAT | GGT  | 'GGC'        | CTT  | TGG     | CGT    | GGC      | CAG   | GCA  | AGG      | GAT   | CCT     | TAGG       |  |  |
| L  |  | L         |        |              |        |         |     |      | Α            |      |         |        |          |       |      |          |       | L       |            |  |  |
|  |  | 25        |        |              |        |         |     |      | 255          | -    |         |        |          |       | _    | 570      |       |         |            |  |  |
| CA   | GAA  | TGA       | GCA    | GCG          | CTG    | GAG     | GTG | GAI  | TTA          | CCG  | TTC     | GGT    | CAT      | CTA   | CGA  | GCC      | CTA   | CCT     | GGCC       |  |  |
| Q  | N  | E         | Q      | R            | W      | R       | W   | I    | F            | Ŗ    | S       | V      | I        | Y     | E    | P        | Y     | L       | A          |  |  |
|  |  | 25        |        |              |        |         |     |      | 261          | -    |         |        |          |       | _    | 630      |       |         |            |  |  |
| ATGTTCGGCCAGGTGCCCAGTGACGTGGATGGTACCACGTATGACTTTGCCCACTGCACC                                       |  |           |        |              |        |         |     |      |              |      |         |        |          |       | CACC |          |       |         |            |  |  |
| M  | F  | G         | Q      | V            | P      | S       | D   | V    | D            | G    | T       | T      | Y        | D     | F    | A        | H     | С       | T          |  |  |
|  | 2650 2670 2690   |           |        |              |        |         |     |      |              |      |         |        |          |       |      |          |       |         |            |  |  |
| TT   | CAC  | TGG       | GAA'   | TGA          | GTC    | CAA     | GCC | AC'I | GTG          | TGT  | GGA     | GCI    | GGA      | TGA   |      |          |       |         | CCGG       |  |  |
| F  | T  | G         | N      | E            | S      | ĸ       | P   | L    | C            | V    | E       | L      | D        | E     | H    | N        | L     | P       | R          |  |  |
|  |  | 27        |        |              |        |         |     |      | 273          |      |         |        |          |       | _    | 750      |       |         |            |  |  |
| TT   | TTCCCCGAGTGGATCACCATCCCCCTGGTGTGCATCTACATGTTATCCACCAACATCCTG |           |        |              |        |         |     |      |              |      |         |        |          |       |      |          |       |         |            |  |  |
| F  | P  | E         | W      | Ι            | T      | I       | ₽   | L    | V            | C    | I       | Y      | M        | · L   | _    | ${f T}$  | N     | I       | L          |  |  |
|  |  | 27        |        |              |        |         |     |      | 279          | -    |         |        |          |       | _    | 810      |       |         |            |  |  |
| CI   | GGT  | CAA       | CCT    |              |        |         |     |      |              |      |         |        |          |       |      |          |       |         | CAAT       |  |  |
| L  | V  | N         | L      | L            | V      | A       | M   | F    | G            |      | ${f T}$ | V      | G        | T     | -    | Q        | E     | N       | N          |  |  |
|  |  | 28        |        |              |        |         |     |      | 285          |      |         |        |          |       | _    | 870      |       |         |            |  |  |
| GΑ   | CCA  | GGT       | CTG    |              |        |         |     |      |              |      |         |        |          |       |      |          |       |         | CAAT       |  |  |
| D  | Q  | V         |        | K            | F      | Q       | R   |      | F            |      | V       | Q      | Е        | Y     |      |          | R     | L       | N          |  |  |
|  |  | 28        |        |              |        |         |     |      | 291          |      |         |        |          |       | _    | 930      | ~     |         |            |  |  |
|  |  |           |        |              |        | CGT     | CTI | CGC  | CTTA         | CTI  | CTA     | CAI    | GGT      | GGI   | GAA  | .GAA     | GIG   | CTT     | CAAG       |  |  |
| Ι  | P  | _         | P      | F            | I      | V       | F   | Α    | Y            |      | Y       | M      | V        | V     |      |          | C     | F.      | K          |  |  |
|  |  | 29        |        |              |        | ·       |     |      | 297          |      |         |        | . ~      |       | _    | 990      | ma »  | 707     | CIN A MI   |  |  |
|  |  |           |        |              |        |         |     |      |              |      |         |        |          |       |      |          |       |         | CAAT       |  |  |
| С  | С  | _         | K      | E            | K      | N       | M   | E    | S            |      | ٧       | C      | Ç        | F.    |      | и<br>050 | Ľ,    | D       | IN         |  |  |
|  |  | 30        |        |              |        |         |     |      | 303          |      |         | ~m>    | 0011     | mon   | _    |          | ת איי | CNC     | תתתתי      |  |  |
|  |  |           |        |              |        |         |     |      |              |      |         |        | L<br>L   | TG1   | K    |          | N     | T       | 'AAAA<br>K |  |  |
| E  | Т  | L         |        | M            | E      | G       | ٧   | ΙM   | К<br>309     | _    | 1/4     | 1      | ш        | V     |      | 110      |       | 1       | 10         |  |  |
| ~  | 107 N  | 30        |        | ama          | מים מו | ~~x     | አአባ | יראנ |              |      | יייי על | מידים. | יא ריז   | אריז  | _    |          |       | ינינייו | TAAT       |  |  |
|  |  |           |        |              |        |         |     |      | JGCA<br>H    |      |         |        |          |       |      |          | K     | L       |            |  |  |
| A  | N  | D<br>31   | _      | s            | E      | E       | 145 | ĸ    | 315          |      | r       | А      | V        | ש     | ע    | -        | **    | _       | ^-         |  |  |
| <b>~</b>   | THAT THE   |           |        | <b>т</b> С10 | 17 CIT | ע ע די. | אמא | ימטי | o ro<br>TTGC |      | י מידי  | יי ע ע | ית מייזי | מידים | ı.G. |          |       |         |            |  |  |
| D  |  |           | G<br>G |              |        |         |     |      | A<br>A       |      | K       |        | K        |       |      |          |       |         |            |  |  |
| ע  | п  | T.        | G      | יו           | יד     | 1/      | E   | -    | A            | T.A. | 10      | -      | 10       |       |      |          |       |         |            |  |  |

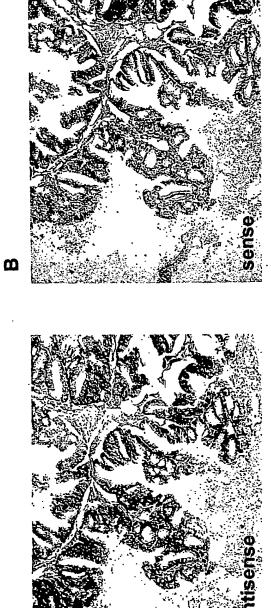
## b.) Trp10 protein:

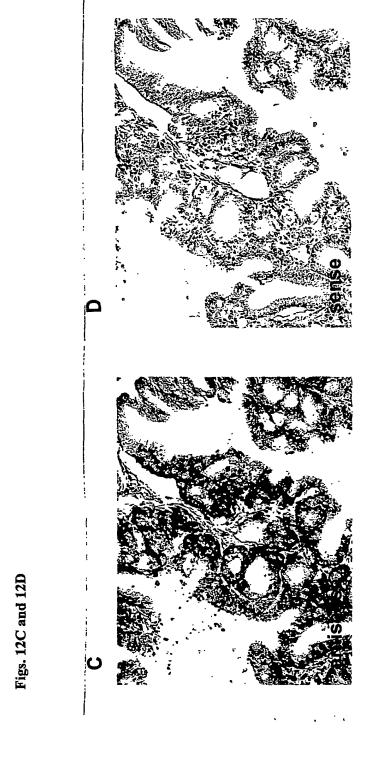
MKSFLPVHTIVLIRENVCKCGYAQSQHMEGTQINQSEKWNYKKHTKEFPTDAFGDIQFETLGKKGKYIRLSCDTDAEILY
ELLTQHWHLKTPNLVISVTGGAKNFALKPRMRKIFSRLIYIAQSKGAWILTGGTHYGLMKYIGEVVRDNTISRSSEENIV
AIGIAAWGMVSNRDTLIRNCDAEGYFLAQYLMDDFTRDPLYILDNNHTHLLLVDNGCHGHPTVEAKLRNQLEKYISERTI
QDSNYGGKIPIVCFAQGGGKETLKAINTSIKNKIPCVVVEGSGQIADVIASLVEVEDALTSSAVKEKLVRFLPRTVSRLP
EEETESWIKWLKEILECSHLLTVIKMEEAGDEIVSNAISYALYKAFSTSEQDKDNWNGQLKLLLEWNQLDLANDEIFTND
RRWESADLQEVMFTALIKDRPKFVRLFLENGLNLRKFLTHDVLTELFSNHFSTLVYRNLQIAKNSYNDALLTFVWKLVAN
FRRGFRKEDRNGRDEMDIELHDVSPITRHPLQALFIWAILQNKKELSKVIWEQTRGCTLAALGASKLLKTLAKVKNDINA
AGESEELANEYETRAVELFTECYSSDEDLAEQLLVYSCEAWGGSNCLELAVEATDQHFIAQPGVQNFLSKQWYGEISRDT
KNWKIILCLFIIPLVGCGFVSFRKKPVDKHKKLLWYYVAFFTSPFVVFSWNVVFYIAFLLLFAYVLLMDFHSVPHPPELV
LYSLVFVLFCDEVRQWYVNGVNYFTDLWNVMDTLGLFYFIAGIVFRLHSSNKSSLYSGRVIFCLDYIIFTLRLIHIFTVS
RNLGPKIIMLQRMLIDVFFFLFLFAVWMVAFGVARQGILRQNEQRWRWIFRSVIYEPYLAMFGQVPSDVDGTTYDFAHCT
FTGNESKPLCVELDEHNLPRFPEWITIPLVCIYMLSTNILLVNLLVAMFGYTVGTVQENNDQVWKFQRYFLVQEYCSRLN
IPFPFIVFAYFYMVVKKCFKCCCKEKNMESSVCCFKNEDNETLAWEGVMKENYLVKINTKANDTSEEMRHRFRQLDTKLN
DLKGLLKEIANKIK

Figs. 12A and 12B

The Trp8 gene is expressed in endometrial or uterine cancer, but not in normal endometrium

**Endometrial cancer:** 





**SUBSTITUTE SHEET (RULE 26)** 

F Springer of the state of the

Figs. 12E and 12F

Endometrium:

SUBSTITUTE SHEET (RULE 26)

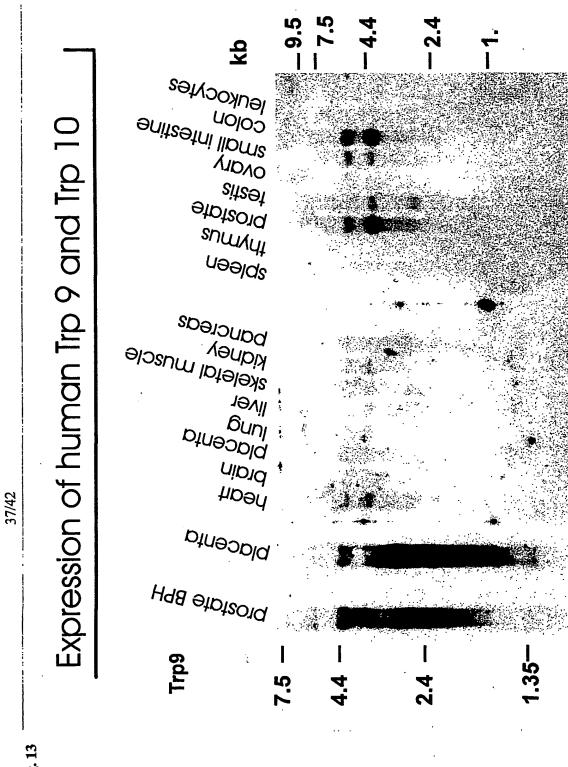
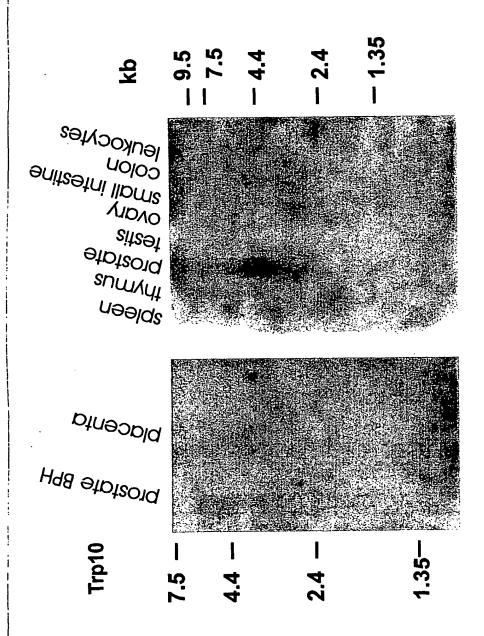
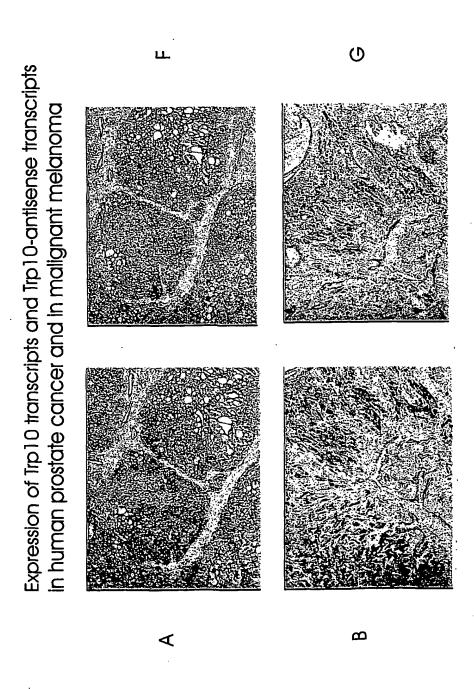


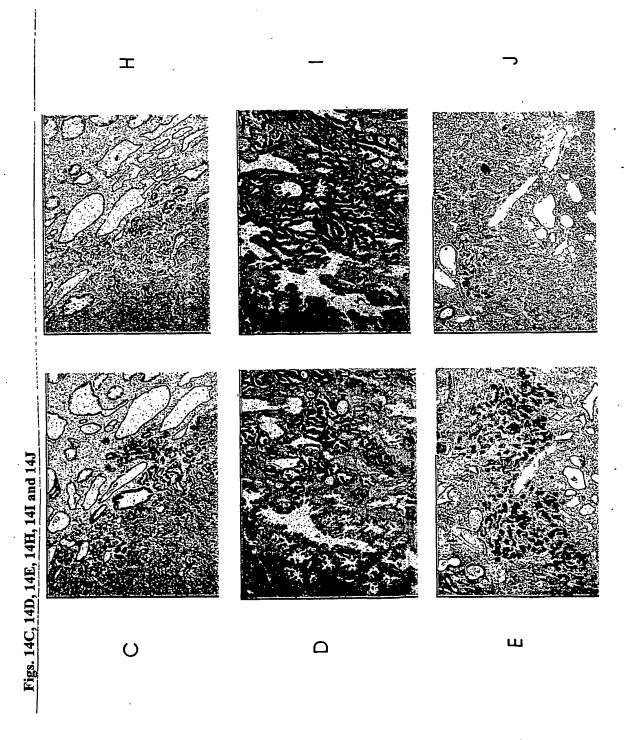
Fig. 13



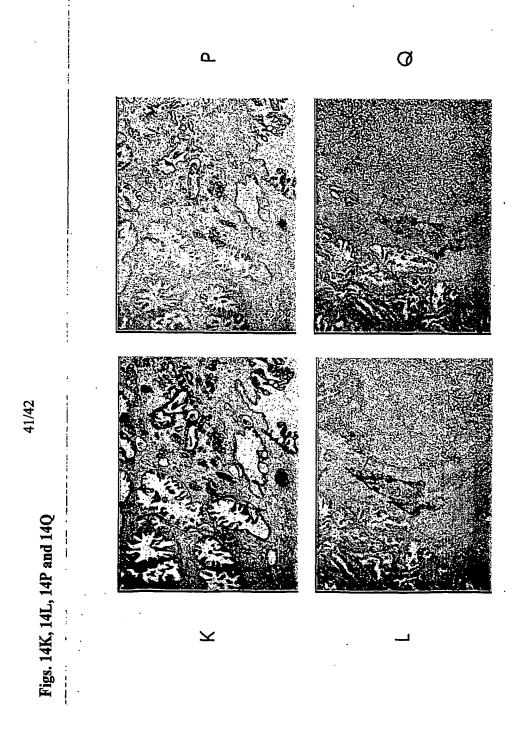


Figs. 14A, 14B, 14F and 14G



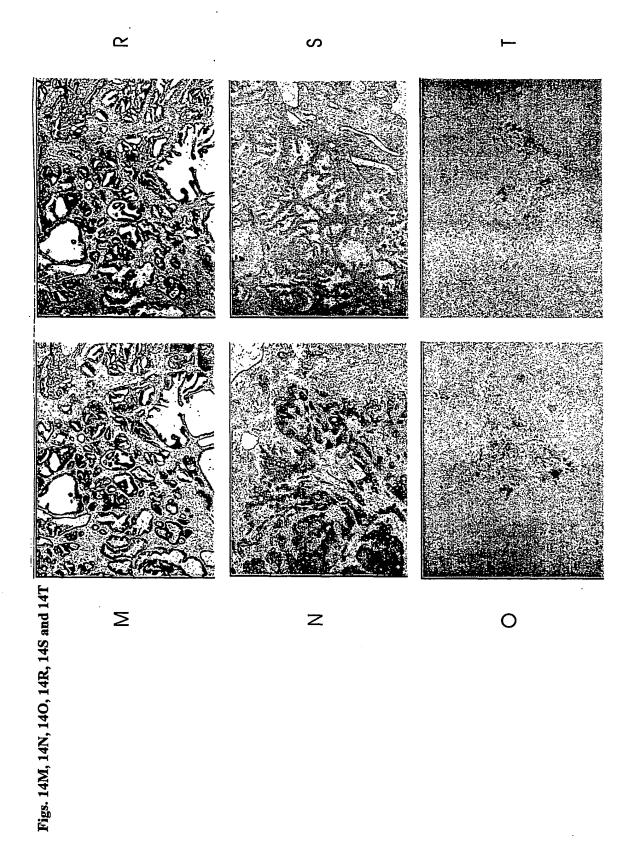


SUBSTITUTE SHEET (RULE 26)



SUBSTITUTE SHEET (RULE 26)

PCT/EP01/08309



## This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

BLACK BORDERS

IMAGE CUT OFF AT TOP, BOTTOM OR SIDES

FADED TEXT OR DRAWING

BLURRED OR ILLEGIBLE TEXT OR DRAWING

SKEWED/SLANTED IMAGES

COLOR OR BLACK AND WHITE PHOTOGRAPHS

GRAY SCALE DOCUMENTS

LINES OR MARKS ON ORIGINAL DOCUMENT

REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY

OTHER:

## IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.